

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 1, 2003, 18:45:51 ; Search time 7.51496 Seconds
(without alignments)
827.642 Million cell updates/sec

Title: US-09-856-927-4
Perfect score: 769
Sequence: 1 FGLGAEAYTASSMALAIATG.....MIIFLTAYLKLFKKYS 147

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/taa/5A-COMB.pep:*
2: /cgn2_6/ptodata/1/taa/5B-COMB.pep:*
3: /cgn2_6/ptodata/1/taa/6A-COMB.pep:*
4: /cgn2_6/ptodata/1/taa/6B-COMB.pep:*
5: /cgn2_6/ptodata/1/taa/PTUS-COMB.pep:*
6: /cgn2_6/ptodata/1/taa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	623	81.0	655	US-09-245-808-1	Sequence 1, Appli
2	623	81.0	655	US-09-767-594-1	Sequence 1, Appli
3	76.5	9.9	482	US-09-107-532A-4868	Sequence 4868, Ap
4	74	9.6	831	US-08-677-734A-11	Sequence 11, Appl
5	74	9.6	831	US-09-097-053-11	Sequence 11, Appl
6	72.5	9.4	266	US-09-194-146-2	Sequence 2, Appli
7	72	9.4	551	US-09-328-352-7176	Sequence 7176, Ap
8	71.5	9.3	641	US-09-687-538B-8	Sequence 8, Appli
9	71	9.2	505	US-09-328-352-4724	Sequence 4724, Ap
10	70.5	9.2	153	US-08-695-736-4	Sequence 4, Appli
11	70.5	9.2	638	US-09-687-538B-4	Sequence 4, Appli
12	70	9.1	375	US-09-252-991A-27411	Sequence 27411, A
13	69.5	9.0	857	US-09-107-532A-4201	Sequence 4201, Ap
14	69	9.0	664	US-09-621-816B-10	Sequence 10, Appl
15	68	8.8	468	US-09-252-991A-22874	Sequence 22874, A
16	68	8.8	483	US-09-107-532A-4123	Sequence 4123, Ap
17	67.5	8.8	155	US-09-107-532A-5926	Sequence 5926, Ap
18	67.5	8.8	466	US-09-328-352-7117	Sequence 7117, Ap
19	67	8.7	907	US-09-198-452A-306	Sequence 306, App
20	66.5	8.6	361	US-08-415-751-36	Sequence 36, Appl
21	66.5	8.6	420	US-09-134-001C-3805	Sequence 3805, Ap
22	66.5	8.6	668	US-09-134-001C-3430	Sequence 3430, Ap
23	66	8.6	390	US-09-328-352-4891	Sequence 4891, Ap
24	66	8.6	617	US-07-879-617A-11	Sequence 11, Appl
25	66	8.6	617	US-08-301-722A-3	Sequence 3, Appli
26	66	8.6	617	US-08-240-783B-3	Sequence 3, Appli
27	66	8.6	617	US-08-753-985-11	Sequence 11, Appl

28	66	8.6	617	3	US-09-084-813-3	Sequence 3, Appli
29	66	8.6	617	5	PCT-US92-09662-3	Sequence 3, Appli
30	66	8.6	901	4	US-09-134-001C-5389	Sequence 5389, Ap
31	65.5	8.5	423	2	US-08-494-907-14	Sequence 14, Appl
32	65.5	8.5	423	5	PCT-US96-10986-14	Sequence 14, Appl
33	65.5	8.5	834	2	US-08-677-734A-9	Sequence 9, Appli
34	65.5	8.5	834	2	US-08-677-734A-10	Sequence 10, Appli
35	65.5	8.5	834	4	US-09-097-053-9	Sequence 9, Appli
36	65.5	8.5	834	4	US-09-097-053-10	Sequence 10, Appli
37	65	8.5	104	1	US-08-111-939-26	Sequence 26, Appl
38	65	8.5	408	4	US-09-107-532A-6198	Sequence 6198, Ap
39	65	8.5	519	4	US-09-328-352-7681	Sequence 7681, Ap
40	65	8.5	875	4	US-09-585-858-18	Sequence 18, Appl
41	65	8.5	1220	2	US-08-680-326-38	Sequence 38, Appl
42	65	8.5	1287	4	US-09-252-991A-29606	Sequence 29606, A
43	65	8.5	2516	3	US-08-374-077C-2	Sequence 2, Appli
44	65	8.5	2516	3	US-08-895-590-2	Sequence 2, Appli
45	65	8.5	2516	4	US-09-533-879A-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1

US-09-245-808-1
; Sequence 1, Application US/09245808
; Patent No. 6313277
; GENERAL INFORMATION:
; APPLICANT: Doyle, L. Austin
; APPLICANT: Abruzzo, Lynne V.
; APPLICANT: Ross, Douglas D.
; TITLE OF INVENTION: Breast Cancer Resistance Protein (BCRP) and DNA which
; FILE OF INVENTION: encodes it
; FILE REFERENCE: Ross Umb conversion
; CURRENT APPLICATION NUMBER: US/09/245,808
; CURRENT FILING DATE: 1999-02-05
; EARLIER APPLICATION NUMBER: 60/073763
; EARLIER FILING DATE: 1998-02-05
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Human MCF-7/AdrVp cells
US-09-245-808-1

Query Match	81.0%	Score 623;	DB 4;	Length 655;
Best Local Similarity	81.6%	Pred. No. 1.9e-62;		
Matches 120;	Conservative	8;	Mismatches 17;	Indels 2;
Gaps	1;			
Qy	1	FGLGAEAYTASSMALAIATGOSVSVATLLMTIAFVFMFLPSGLLVNLTIGPWLWSLQY	60	
Db	511	FTLMVAVSASSMALAIAGSVSVATLLMTICFVFMFLPSGLLVNLTIASWLSLQY	570	
Qy	61	FSIPRYGFTALQYNEFLGQEPFGPNVTDNSTCVNSYAICTGNEYLINQGIELSPWGLWK	120	
Db	571	FSIPRYGFTALQYNEFLGQEPFGPNVTDNSTCVNSYAICTGNEYLINQGIELSPWGLWK	628	
Qy	121	NHVALACMIIFLTAYLKLFKKYS	147	
Db	629	NHVALACMIIFLTAYLKLFKKYS	655	

RESULT 2

US-09-767-594-1
; Sequence 1, Application US/09767594
; Patent No. 6521635
; GENERAL INFORMATION:
; APPLICANT: Rates, Susan
; APPLICANT: Robey, Robert
; APPLICANT: The Government of the United States of America
; APPLICANT: as represented by the Secretary of the
; APPLICANT: Department of Health and Human Services

;; TITLE OF INVENTION: Inhibition of MXR Transport by Acridine Derivatives
;; FILE REFERENCE: 015280-402100US
;; CURRENT APPLICATION NUMBER: US/09/767,594
;; CURRENT FILING DATE: 2001-01-22
;; PRIOR APPLICATION NUMBER: US 60/177,410
;; PRIOR FILING DATE: 2000-01-20
;; NUMBER OF SEQ ID NOS: 2
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 1
;; LENGTH: 655
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: human mitoxanthrone resistance (MXR)/BRCP/ABCP
US-09-767-594-1

Query Match 81.0%; Score 623; DB 4; Length 655;
Best Local Similarity 81.6%; Pred. No. 1.9e-62;
Matches 120; Conservative 8; Mismatches 17; Indels 2; Gaps 1;
QY 1 FGLGAEAYTASSMALAIATGQSVSVATLLMTIAFVFMMLFSGLLVNLRTIGPWSLWLOY 60
DB 511 FTLMMVAYSASSMALAIAGQSVSVATLLMTICFVFMNIFSGLLVNLRTIASLWLOY 570
QY 61 FSPRYGFTALQYNEFLGQFCPGENVNTDNCVNSYAICTGNEYLINQIELSPWGLWK 120
DB 571 FSPRYGFTALQYNEFLGQFCPGENVNTDNCVNSYAICTGNEYLINQIELSPWGLWK 120
QY 121 NHVALACMIITFLTIAYLKLFLKKYS 147
DB 629 NHVALACMIIVFTIAYLKLFLKKYS 655

RESULT 3
US-09-107-532A-4868
; Sequence 4868, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 4868:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 482 amino acids

;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; HYPOTHETICAL: YES
;; ORIGINAL SOURCE:
;; ORGANISM: Enterococcus faecium
;; FEATURE:
;; NAME/KEY: misc feature
;; LOCATION: (B) LOCATION 1...482
;; SEQUENCE DESCRIPTION: SEQ ID NO: 4868:
US-09-107-532A-4868

Query Match 9.9%; Score 76.5; DB 4; Length 482;
Best Local Similarity 24.1%; Pred. No. 1.9;
Matches 41; Conservative 19; Mismatches 45; Indels 65; Gaps 10;
QY 3 LGAEAYTASSMALAIATGQSVSVATLLMTIAF--VFMMLF-----SGLLVNL--RT 50
DB 27 MGLGTFIGLTWALC-ATVRSIPTLAAGVMTLIFYFIFAVLPFAGPISMVSGELSTMLPQE 85
QY 51 IGP-----WLSWQYFSIPRYGFTALQYNEFLGQFCPGENVNTDNCSTC 93
DB 86 GGPQLMVKATLGSKWGFVAMLLWYQMF--PGMVVASTLGLLGNTP----- 131
QY 94 VNSYAICTGNEYLINQIELSPWGLWKNH-VALACMIITFLTIAYLKLFL 142
DB 132 -----GN-----VELG-----NNHWFVLGCILVIYWIITILNLKF 161

RESULT 4
US-08-677-734A-11
; Sequence 11, Application US/08677734A
; Patent No. 5871919
; GENERAL INFORMATION:
; APPLICANT: Brant, Steven R.
; APPLICANT: Yun, Chris C.H.
; APPLICANT: Donowitz, Mark
; APPLICANT: Tee, Chung-Ming
; TITLE OF INVENTION: Cloning, Tissue Distribution, and
; TITLE OF INVENTION: Functional Analysis Of The Human Na+/H+ Exchanger Isoform,
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/677,734A
; FILING DATE: 10-JUL-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fordis, Jean B.
; REGISTRATION NUMBER: 32,984
; REFERENCE/DOCKET NUMBER: 05387.0043-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 831 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide

Db 222 WLVQSKYPOQLHYFFIDQPNRFSKEF-----NNK----- 254
Qy 112 ELSPGLKKNHVALACMIIFLTIAVLKLLFLKKYS 147
Db 255 --QPW-----CFYLMILFVSFLPWLFSRFT 278

RESULT 8
US-09-687-5388-8
; Sequence 8, Application US/09687538B
; Patent No. 6514739
; GENERAL INFORMATION:
; APPLICANT: Udagawa, Hiroaki
; APPLICANT: Frandsen, Torben
; APPLICANT: Nielsen, Tom
; APPLICANT: Kauppinen, Markus
; APPLICANT: Christensen, Soeren
; TITLE OF INVENTION: Lysophospholipase
; FILE REFERENCE: 5958.210-US
; CURRENT APPLICATION NUMBER: US/09/687,538B
; CURRENT FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 8
; LENGTH: 641
; TYPE: PRT
; ORGANISM: Aspergillus oryzae
US-09-687-5388-8

Query Match 9.3%; Score 71.5; DB 4; Length 641;
Best Local Similarity 27.7%; Pred. No. 10;
Matches 26; Conservative 16; Mismatches 45; Indels 7; Gaps 4;
Qy 4 GAERYTASSMALAIATGQSVSVATLLMTIAFVFMMLFSGLLVNLRTIGPWLSWLYQFSI 63
Db 257 GGPSTYSSIALTDTDFKQADMPMLVADGRYPDELWSS-NATVYEFNPW--EFGTFDP 313
Qy 64 PRYGTALQNEFLGQFCFGNVTDNSTCVNSY 97
Db 314 TVYGFVPLEY---VGSKF-DGGSIPDNCTCRGF 343

RESULT 9
US-09-328-352-4724
; Sequence 4724, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4724
; LENGTH: 505
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-4724

Query Match 9.2%; Score 71; DB 4; Length 505;
Best Local Similarity 24.3%; Pred. No. 8.3;
Matches 41; Conservative 33; Mismatches 63; Indels 32; Gaps 8;
Qy 4 GAERYTASSMALA-----IATGQSVSVATLLMTIAFVFMMLFSGLL-----V 46
Db 244 GAPAPMATELATAKVAATVGLFVRYLLASGAIMVNSLVTLTIAVLISLVGNLLAVRQV 303
Qy 47 NLRTIGPWSLYQFSIPRYGF--TALQYNEFLGQFCFGNVTDNSTCVNSY---AICT 101
Eb 304 NUKRI-----LGYSIAHFGYLLIALISMTYASLSGVSVTYVYVTVLTITIGAFGVALMS 357

Qy 102 GNEYLINQIEIISPW-GL-WKNHVALACMIIFLTIAVLKLL--FLKKY 146
Db 358 SPYNNVDEAQSADRYGLFWRRPVLTAITLVWMLSLAGIPLTAGFIGKF 406
RESULT 10
US-08-695-736-4
; Sequence 4, Application US/08695736
; Patent No. 5843714
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Bandman, Olga
; APPLICANT: Goli, Surya K.
; APPLICANT: Hillman, Jennifer L.
; TITLE OF INVENTION: NOVEL HUMAN PROTEOLIPID
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U.S.
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/695,736
; FILING DATE: Filed Herewith
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0108 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 153 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 126719
US-08-695-736-4

Query Match 9.2%; Score 70.5; DB 2; Length 153;
Best Local Similarity 33.3%; Pred. No. 2;
Matches 20; Conservative 11; Mismatches 20; Indels 9; Gaps 3;
Qy 13 MALAIATGQSVV---SVATLLMTIAFVFMMLFSGLL--LVNLRTIGPW---LSWLYQFSI 63
Db 1 MAPAAATGSGTLPSCFSVFTTLPDLLFFIFRIFGLVWLVASSIAPNPVLQGWWMFYSV 60
RESULT 11
US-09-687-5388-4
; Sequence 4, Application US/09687538B
; Patent No. 6514739
; GENERAL INFORMATION:
; APPLICANT: Udagawa, Hiroaki
; APPLICANT: Frandsen, Torben
; APPLICANT: Nielsen, Tom
; APPLICANT: Kauppinen, Markus
; APPLICANT: Christensen, Soeren
; TITLE OF INVENTION: Lysophospholipase
; FILE REFERENCE: 5958.210-US
; CURRENT APPLICATION NUMBER: US/09/687,538B
; CURRENT FILING DATE: 1999-10-20

CORRESPONDENCE ADDRESS:
ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD/ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 4201:
SEQUENCE CHARACTERISTICS:
LENGTH: 857 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...857
SEQUENCE DESCRIPTION: SEQ ID NO: 4201:
US-09-107-532A-4201
Query Match 9.0%; Score 69.5; DB 4; Length 857;
Best Local Similarity 23.5%; Pred. No. 25;
Matches 42; Conservative 25; Mismatches 51; Indels 61; Gaps 9
Qy 11 SSNALATATQGVV-----SVATLLMTIAFVFM-LFSGI-----LVNLETI 51
Db 633 SRPEVVAEGRVVNNIERSASLFLVKNIFSLSLFSVFALTYPLEPSQITLISLFTI 692
Qy 52 GPWLSMLQYFSIPRYGPTALQYNE-----PLGQEFQCPG----- 84
Db 693 G-----LPSF-LLALEENKRRIGKFINNMEKAVPGGTDIMVGVALVCGVTL 741
Qy 85 -FNVTDNSTCVNSYAICTGNETYGLINQGIELSPGLMKNHVALACMIIFLTIAYLKLF 142
Db 742 DLNKTDVSTASTMLLIAVG--FLVLYKI-CSPLNKRFSQIILFCASGIFPSVVFHLKLF 797
RESULT 14
US-09-621-816B-10
Sequence 10, Application US/09621816B
Patent No. 6518488
GENERAL INFORMATION:
APPLICANT: Agarwal, Ameeta
APPLICANT: Liu, Jingdong
APPLICANT: Lahiri, Devlina
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated
with the Beta-Oxidation Pathway
FILE REFERENCE: 16517.071
CURRENT APPLICATION NUMBER: US/09/621,816B
CURRENT FILING DATE: 2002-05-01
NUMBER OF SEQ ID NOS: 10

```

; SEQ ID NO 10
; LENGTH: 664
; TYPE: PRT
; ORGANISM: Glycine max
US-09-621-816B-10

Query Match      9.0%; Score 69; DB 4; Length 664;
Best Local Similarity 23.9%; Pred. No. 20;
Matches 34; Conservative 13; Mismatches 39; Indels 56; Gaps 6;

QY 5 AEAYTASSMALAIATGQSVV-----SVATLLMTTIAFVPMFLFSGLLVNLRTI 51
Db 290 ADASVALSRVAVCIATRYSAVRQFGSHNGGLETOVDYKTQOARLPPLASAYA--FRV 347
QY 52 GPWLSWLQYFSIPRYGFTALQYNERLQBFQFCGFNVTDNSTCVNSYA----- 98
Db 348 GGWLKWL-YNQVTE-----RLQANDF-----STLPEAHACTAGLKSLTTTAT 388
QY 99 -----ICTGNEYLINQGI 111
Db 389 ADGIECRKLCGGHGYLCSSGL 410

RESULT 15
US-09-252-991A-22874
; Sequence 22874, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 22874
; LENGTH: 468
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-22874

Query Match      8.8%; Score 68; DB 4; Length 468;
Best Local Similarity 39.7%; Pred. No. 17;
Matches 23; Conservative 8; Mismatches 25; Indels 2; Gaps 2;

QY 2 GLGAEAYTASSMALAIATGQSVVSVATLLMTTIAFVPMFLFSGLLVNLRTIGP-WLSWL 58
Db 389 GPAELATASTVVGASAIGLPVSTHTLVGAVLIGMARGIGAL-NLRVIGSIFLSWV 445

Search completed: August 1, 2003, 18:52:41
Job time : 9.51496 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 1, 2003, 18:45:51 ; Search time 33.485 Seconds
(without alignments)
827.642 Million cell updates/sec

Title: US-09-856-927-2

Perfect score: 3350

Sequence: 1 MSSNVEVFPVSGNTNGF.....MIVFLTIAYLKLFLKKYS 655

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

- 1: /cgn2_6/ptodata/1/1aa/5A-COMB.pep.*
- 2: /cgn2_6/ptodata/1/1aa/5B-COMB.pep.*
- 3: /cgn2_6/ptodata/1/1aa/6A-COMB.pep.*
- 4: /cgn2_6/ptodata/1/1aa/6B-COMB.pep.*
- 5: /cgn2_6/ptodata/1/1aa/PCTUS-COMB.pep.*
- 6: /cgn2_6/ptodata/1/1aa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3350	100.0	655	US-09-767-594-1	Sequence 1, Appli
2	3331	99.4	655	US-09-245-808-1	Sequence 1, Appli
3	241	7.2	270	US-09-134-001C-5561	Sequence 5561, Ap
4	236.5	7.1	1684	US-08-665-259-25	Sequence 25, Appl
5	236.5	7.1	1684	US-08-762-500-25	Sequence 25, Appl
6	236.5	7.1	1704	US-08-762-500-75	Sequence 75, Appl
7	229.5	6.9	273	US-09-134-001C-3641	Sequence 3641, Ap
8	229	6.8	361	US-09-107-532A-4490	Sequence 4490, Ap
9	227	6.7	256	US-09-134-001C-4600	Sequence 4600, Ap
10	225.5	6.7	382	US-09-134-001C-3439	Sequence 3439, Ap
11	224.5	6.7	1302	US-08-232-537-2	Sequence 2, Appli
12	222.5	6.6	3405	US-09-107-532A-3849	Sequence 3849, Ap
13	219	6.5	1349	US-08-612-7348-2	Sequence 2, Appli
14	216	6.4	329	US-09-107-532A-4844	Sequence 4844, Ap
15	213.5	6.4	365	US-09-328-352-7027	Sequence 7027, Ap
16	213	6.4	1408	US-08-612-521-2	Sequence 2, Appli
17	212	6.3	402	US-09-107-532A-5360	Sequence 5360, Ap
18	211.5	6.3	322	US-09-107-532A-4662	Sequence 4662, Ap
19	211.5	6.3	1457	US-08-665-259-27	Sequence 27, Appl
20	211.5	6.3	1457	US-08-762-500-27	Sequence 27, Appl
21	210	6.3	111	US-08-466-886-31	Sequence 31, Appl
22	210	6.3	111	US-08-466-617-31	Sequence 31, Appl
23	210	6.3	330	US-09-252-991A-27569	Sequence 27569, A
24	208	6.2	240	US-09-198-452A-769	Sequence 769, App
25	207	6.2	256	US-09-107-532A-4208	Sequence 4208, Ap
26	206.5	6.2	528	US-09-107-532A-6340	Sequence 6340, Ap
27	205.5	6.1	345	US-09-252-991A-18872	Sequence 18872, A

ALIGNMENTS

RESULT 1

US-09-767-594-1

; Sequence 1, Application US/09767594

; Patent No. 6521635

; GENERAL INFORMATION:

; APPLICANT: Bates, Susan

; APPLICANT: Robey, Robert

; APPLICANT: The Government of the United States of America

; APPLICANT: as represented by the Secretary of the

; APPLICANT: Department of Health and Human Services

; TITLE OF INVENTION: Inhibition of MXR Transport by Acridine Derivatives

; FILE REFERENCE: 015280-402100US

; CURRENT APPLICATION NUMBER: US/09/767,594

; PRIORITY FILING DATE: 2001-01-22

; PRIOR APPLICATION NUMBER: US 60/177,410

; PRIORITY FILING DATE: 2000-01-20

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 1

; LENGTH: 655

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; OTHER INFORMATION: human nitoxanthrone resistance (MXR)/BRCP/ABCP

; OTHER INFORMATION: protein

US-09-767-594-1

Query Match	100.0%;	Score 3350;	DB 4;	Length 655;
Beat Local Similarity	100.0%;	Pred. No. 0;		
Matches 655;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MSSNVEVFPVSGNTNGFPATVSNDLKATGAVLSFHNICYRVKLSKGFPLCKRQVE	60	
Db	1	MSSNVEVFPVSGNTNGFPATVSNDLKATGAVLSFHNICYRVKLSKGFPLCKRQVE	60	
Qy	61	KEILSNINGIMKGNAILGPTGGKSLDVLAAKDPGSLGVDLNGAPRANPKCN	120	
Db	61	KEILSNINGIMKGNAILGPTGGKSLDVLAAKDPGSLGVDLNGAPRANPKCN	120	
Qy	121	SGYVVQDDVNGTITVENLQFSAALRLATMTNHNKNERINRVIEELGDKVADSKVGT	180	
Db	121	SGYVVQDDVNGTITVENLQFSAALRLATMTNHNKNERINRVIEELGDKVADSKVGT	180	
Qy	181	QFIRGVSGGERKRTSIGMELITDPSILSLDPTTGLDSSTANAVLLLLKRMKSKQRTIIF	240	
Db	181	QFIRGVSGGERKRTSIGMELITDPSILSLDPTTGLDSSTANAVLLLLKRMKSKQRTIIF	240	
Qy	241	SIHQPRYSIFKLFDSLTLASGRMLPHGPAQEGALGPESAGYHCEAYNNPADFFLDIING	300	
Db	241	SIHQPRYSIFKLFDSLTLASGRMLPHGPAQEGALGPESAGYHCEAYNNPADFFLDIING	300	

QY 301 DSTAVLNREEDFKATEIIEPSKQDKPLIEKLAIEYVNSSFYKETKAEHLHOLSGGKXXX 360
DB 301 DSTAVLNREEDFKATEIIEPSKQDKPLIEKLAIEYVNSSFYKETKAEHLHOLSGGKXXX 360
QY 361 ITVFKIEISYTTSFCHQLRWVSKRSPKLLGNPOASIAQIIVTVLGLVIGAIYFGLKXDS 420
DB 361 ITVFKIEISYTTSFCHQLRWVSKRSPKLLGNPOASIAQIIVTVLGLVIGAIYFGLKXDS 420
QY 421 TGIQNRAGVLPFLTTNOCFSSVSARELVFVEKKLFIHEYISGYRVSSYFGLKLLSLLP 480
DB 421 TGIQNRAGVLPFLTTNOCFSSVSARELVFVEKKLFIHEYISGYRVSSYFGLKLLSLLP 480
QY 481 MRMLPSIIFTCIVFVFMGLKPKADAFVMMFTLMWVAYSASSMALAIAAGQSVSVATLL 540
DB 481 MRMLPSIIFTCIVFVFMGLKPKADAFVMMFTLMWVAYSASSMALAIAAGQSVSVATLL 540
QY 541 MTICFVFMIFSGLLVNLTTIASWLSWLYFSIPRYGFTALQHNFLGQFCPLNATGN 600
DB 541 MTICFVFMIFSGLLVNLTTIASWLSWLYFSIPRYGFTALQHNFLGQFCPLNATGN 600
QY 601 NPCNYATCTGEEYLVKQIDLSPLWGLWKNHVALACMIVIFLTIAVYKLLFLKYS 655
DB 601 NPCNYATCTGEEYLVKQIDLSPLWGLWKNHVALACMIVIFLTIAVYKLLFLKYS 655

RESULT 2

US-09-245-808-1

; Sequence 1, Application US/09245808

; Patent No. 6313277

; GENERAL INFORMATION:

; APPLICANT: Doyle, L. Austin

; APPLICANT: Abruzzo, Lynne V.

; APPLICANT: Ross, Douglas D.

; TITLE OF INVENTION: Breast Cancer Resistance Protein (BCRP) and DNA which

; FILE REFERENCE: encodes it

; CURRENT APPLICATION NUMBER: US/09/245,808

; PRIOR FILING DATE: 1999-02-05

; EARLIER APPLICATION NUMBER: 60/073763

; EARLIER FILING DATE: 1998-02-05

; NUMBER OF SEQ ID NOS: 7

; SOFTWARE: Patent In Ver. 2.0

; SEQ ID NO 1

; TYPE: PRT

; LENGTH: 655

; ORGANISM: Human MCF-7/AdrVp cells

US-09-245-808-1

Query Match 99.4%; Score 3311; DB 4; Length 655;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 651; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSSNVVEFIPVSQNTNGFPATVNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKPVE 60
DB 1 MSSNVVEFIPVSQNTNGFPATVNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKPVE 60
QY 61 KEILSNINGIMKPGNALIGPTGGCKSSLLDLAARKDPSGLSGDVLINGAPRANFKN 120
DB 61 KEILSNINGIMKPGNALIGPTGGCKSSLLDLAARKDPSGLSGDVLINGAPRANFKN 120
QY 121 SGYVVQDDVMGTLTVRENLOFSAALRLATTMTNHEKNERINRVIEELGLDKVADSKVGT 180
DB 121 SGYVVQDDVMGTLTVRENLOFSAALRLATTMTNHEKNERINRVIEELGLDKVADSKVGT 180
QY 181 QPIRVSGGERKRTSIGMELITDPSILSDDEPTGLDSTANAVALLLKRMSKQRTIIF 240
DB 181 QPIRVSGGERKRTSIGMELITDPSILSDDEPTGLDSTANAVALLLKRMSKQRTIIF 240
QY 241 SIHQPRYSIFKLFDSLTLLASRLMFHGPQAQALCYFESAGVHCAYNNPADFELDIING 300
DB 241 SIHQPRYSIFKLFDSLTLLASRLMFHGPQAQALCYFESAGVHCAYNNPADFELDIING 300

QY 301 DSTAVLNREEDFKATEIIEPSKQDKPLIEKLAIEYVNSSFYKETKAEHLHOLSGGKXXX 360
DB 301 DSTAVLNREEDFKATEIIEPSKQDKPLIEKLAIEYVNSSFYKETKAEHLHOLSGGKXXX 360
QY 361 ITVFKIEISYTTSFCHQLRWVSKRSPKLLGNPOASIAQIIVTVLGLVIGAIYFGLKXDS 420
DB 361 ITVFKIEISYTTSFCHQLRWVSKRSPKLLGNPOASIAQIIVTVLGLVIGAIYFGLKXDS 420
QY 421 TGIQNRAGVLPFLTTNOCFSSVSARELVFVEKKLFIHEYISGYRVSSYFGLKLLSLLP 480
DB 421 TGIQNRAGVLPFLTTNOCFSSVSARELVFVEKKLFIHEYISGYRVSSYFGLKLLSLLP 480
QY 481 MRMLPSIIFTCIVFVFMGLKPKADAFVMMFTLMWVAYSASSMALAIAAGQSVSVATLL 540
DB 481 MRMLPSIIFTCIVFVFMGLKPKADAFVMMFTLMWVAYSASSMALAIAAGQSVSVATLL 540
QY 541 MTICFVFMIFSGLLVNLTTIASWLSWLYFSIPRYGFTALQHNFLGQFCPLNATGN 600
DB 541 MTICFVFMIFSGLLVNLTTIASWLSWLYFSIPRYGFTALQHNFLGQFCPLNATGN 600
QY 601 NPCNYATCTGEEYLVKQIDLSPLWGLWKNHVALACMIVIFLTIAVYKLLFLKYS 655
DB 601 NPCNYATCTGEEYLVKQIDLSPLWGLWKNHVALACMIVIFLTIAVYKLLFLKYS 655

RESULT 3

US-09-134-001C-5561

; Sequence 5561, Application US/09134001C

; Patent No. 6380370

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: GTC-007

; CURRENT APPLICATION NUMBER: US/09/134,001C

; CURRENT FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: US 60/064,964

; PRIOR FILING DATE: 1997-11-08

; PRIOR APPLICATION NUMBER: US 60/055,779

; PRIOR FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5674

; SEQ ID NO 5561

; LENGTH: 270

; TYPE: PRT

; ORGANISM: Staphylococcus epidermidis

US-09-134-001C-5561

Query Match 7.2%; Score 241; DB 4; Length 270;
Best Local Similarity 27.1%; Pred. No. 1.4e-16;
Matches 81; Conservative 55; Mismatches 103; Indels 60; Gaps 12;

QY 42 ICYRVKL-----KSGFLPC-----RKPEKEILSNINGIMKPG-LNALIGPTG 83
DB 1 VCLRRKKIWLKYDRSDYLSILEVRQLSKVYGDQKQAFQEVLEKINLDVDEGEFISIMGSPG 60
QY 84 GGSLSLLDLAARKDPSGLSGDVLING--APRAN-----FKCNSGVVVDVVMGTLT 135
DB 61 SKKTLNLVLSIDYMT--KGSITNGKLEKLSNQLSDIRKDKDIGFIDYINILNTLT 118
QY 136 VRENLOFSAALRLATTMTNHEKNERINRVIEELGLDKVADSKVGTQFIRGVSGGERKRTS 195
DB 119 VKENIMPLSVQKLDKQIMH---ERYQRIVEALNISDISD-----KYPSELSGQORQRTS 170
QY 196 IGMELITDPSILSDDEPTGLDSTANAVALLLKRMSKQ-ORTIIFSIHQPRYSIFKLF 254
DB 171 AARAFINPLSIFADEPTGLDSDKSTLTLKRLKYMNEEFNTTILMTVTHDPVAASFS--N 228
QY 255 SLTLASRLMFHGPQAQALCYFESAGVHCAYNNPAD---FFLDIINGDSTAVLNRE 310
DB 229 RVMLKDGQIF-----TELYQGDGDDDKQTFYKEIIRTQSVLGGIINYE 269

RESULT 4


```
US-08-665-259-25
; Sequence 25, Application US/08665259
; Patent No. 6028173
; GENERAL INFORMATION:
; APPLICANT: Landes, Gregory M.
; APPLICANT: Burn, Timothy C.
; APPLICANT: Connors, Timothy D.
; APPLICANT: Dackowski, William R.
; APPLICANT: Van Raay, Terence J.
; APPLICANT: Klinger, Katherine W.
; TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,
; METHODS OF MAKING AND USING SAME
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENZYME CORPORATION
; STREET: One Mountain Road
; CITY: Framingham
; STATE: Massachusetts
; COUNTRY: United States of America
; ZIP: 01701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 17-JUN-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Dugan, Deborah A.
; REGISTRATION NUMBER: 37,315
; REFERENCE/DOCKET NUMBER: IGS-9.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508) 872-8400
; TELEFAX: (508) 872-5415
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1684 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-665-259-25

Query Match 7.1%; Score 236.5; DB 3; Length 1684;
Best Local Similarity 20.3%; Pred. No. 8.9e-15;
Matches 169; Conservative 116; Mismatches 241; Indels 305; Gaps 39;

QY 66 NINGIMKPLNAILGPTGGKSSLDVLAARKDPSGLSGDVLLG---APRPANPKCNSG 122
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 533 NLN-LYEQITVLLGHNGAGKTTLSMLTGLPPT--SGRAYISGVEISQDMVQIRKSIG 589
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 123 YVODVVMGTLTVRENLOFSALRALATTWTHKERNINRVIELGLDKVADSKVGTFQ 182
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 590 LCPQHDLFDNLTVAEHLFYFAQLK---GLSRQKCPPEVKQMLHIGLEDKWNRS--- 641
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 183 IRGSGGGRKRTSIGMELITDPSILSLDPTTGLDSSSTANAVLLLKRMKSKGRTIIFS 242
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 642 SRPLSGGRRKLSIGIALAGSKVILDEPTSGMDAISRAIWDLLQR-QKSDRIIVLT 700
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 243 HOPRYSIFK-----LPDSLTLIASGRLMFHPGA---QEALGYFESAGYHC----- 290
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 701 H-----FMDEADLLGDRIAINAKGELQCCGSSLFLKQYK---AGYHMTLVKPHCNP 750
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 291 AD-----FFLD----- 296
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 751 EDISOLVHHVHPNATLESSAGAELSFILPRESTRHFEGLFAKLEKKQKELGIASPGASIT 810
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 297 ----- 300
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 811 TMEEVFLRVGLVDSSMDIQALQALQYQHERRASDWDVNSLNCGLMDPSDGICALIEE 870
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 301 DSTAVLN-----REEDFKATEIIEP-----SKQ- 324
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Db 871 ERTAVKLTGLALHCQQFWMAMPLKKAAYSREKWAQVLPVLTCTVTALLAINYSSEL 930
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 325 -DKPLI-----EKLAE-----IYVNSSFYKETAELHQ----- 351
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 931 FDDPMLRLTGEYGRVVPSPVPGTSQGLQQLSEHLKDALQAEQEPREVGLDLEEFILF 990
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 352 ---LSGGEKKKKTIVFKBISYTTSFCHQLRWVSKRSFKNLGNPQ-----ASIAQI 399
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 991 RASVEGGGFNERCLV-----AASF-----RDVGRVTVVALFNQAYHSPATALAVVDNL 1040
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 400 IVTVVLGLVIGAIYFGLKNDSTGIQ-----NRAGVLFFLTNQCF-----SVSAV 445
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1041 LFKLLCGPHASIVVSNFPQPRSAQAKQDQNEGRKGFDIALNLLFAMAFLASTFSLAV 1100
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 446 ELFVVEKKLFTH-EYISGYRVSVYFLKLSLDLPMRLPSIIFTCIVYFMLGLKPKAD 504
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1101 SERAVQAK---HVQFVSGVH-VASFWSALLWDLISP-LIFSLLL--LVVF-----K 1145
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 505 AFFVMFTLMVAYSASSMALAIAGOSVSVATLLMTICFVFMVIFSGLLVNLTTIASW 564
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1146 APDVRAPT--RDGHMADTLLLLLYGWAIIPLMVLMN-----FFPLGAATAYTRLTIFNI 1198
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 565 LSWLOYP-----SIPRYGP-----TALOHNEFLGQFCPLNATGNNFCNYAT---CTGE 611
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1199 LSGIATFLMVTIMRIPAVKLELSKTLDHVFLVLPNHCLGM-AVSSFYENYETRYCTSS 1257
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 612 E-----YLVKQIDILSPW---GLWKNHVALACMIVIFLTIAYLKLLFL 651
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1258 EVAHYCKYNIQENFYAWSAPGVGRFVASMAAS-----GCAYLILLFL 1303
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 5
US-08-762-500-25
; Sequence 25, Application US/08762500
; Patent No. 6030806
; GENERAL INFORMATION:
; APPLICANT: Landes, Gregory M.
; APPLICANT: Burn, Timothy C.
; APPLICANT: Connors, Timothy D.
; APPLICANT: Dackowski, William R.
; APPLICANT: Van Raay, Terence J.
; APPLICANT: Klinger, Katherine W.
; TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,
; METHODS OF MAKING AND USING SAME
; NUMBER OF SEQUENCES: 83
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENZYME CORPORATION
; STREET: One Mountain Road
; CITY: Framingham
; STATE: Massachusetts
; COUNTRY: United States of America
; ZIP: 01701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 09-DEC-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/762,500
; FILING DATE: 17-JUN-1996
; APPLICATION NUMBER: US 08/665,259
; FILING DATE: 17-JUN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/10469
; FILING DATE: 17-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Dugan, Deborah A.
; REGISTRATION NUMBER: 37,315
; REFERENCE/DOCKET NUMBER: IGS-9.3
; TELECOMMUNICATION INFORMATION:
```

TELEPHONE: (508) 872-8400
TELEFAX: (508) 872-5415
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 1684 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-762-500-25

Query Match 7.1%; Score 236.5; DB 3; Length 1684;
Best Local Similarity 20.3%; Pred. No. 8.9e-15;
Matches 169; Conservative 116; Mismatches 241; Indels 305; Gaps 39;

QY 66 NINGIMKPGNLAILGPTGGKSSLDVLAARKDPSGLSGDVLING---APRPANFKNSG 122
DB 533 NLN-LYEGQITVLLGHNGAGKTTTSLMLTGLFPPT--SGRAVISGEYSQDMVQIRKSLG 589
QY 123 YVQDDVVMGTLTVRENLOFSAALRLATTMTNHEKNERINRVIEELGLDKVADSKVGTQF 182
DB 590 LCPQHDILFDNLTVAEHLVYFAQLK---GLSRQKCPPEVKQMLHIIGLEDKWSR-----641
QY 183 IRGVSGGERKRTSIGMELITDPSILSLDEPTTGLSDSTANAVLLLLKRMKSGQRTIIFSI 242
DB 642 SRFLSGMRKLSIGIALIAGSKVLIDLEPTSGMDAISRAIWDLLQR-QKSDRIIVLT 700
QY 243 HQPRYSIFK---LFDSTLLASGLMHPGA---QEALGYFESAGYHC-----BAYNPP 290
DB 701 H-----FMDEADLLGDRIAMWAGELQCCGSSFLKQYK---AGYHMTLVKPEHCNP 750
QY 291 AD-----FFLD-----296
DB 751 EDISQVHHVHPNATLESSAGAEISFILPRESTRHFEGLFAKLEKKQKELGIASFASIT 810
QY 297 -----ING 300
DB 811 TMEEVFLRVGLVDSSMDIQALPALQYQHERRASDWAVDNLCGAMPDSDGICALIEE 870
QY 301 DSTAVALN-----REEDFKATILIEP-----SKQ- 324
DB 871 ERTAVKLNTGLALHCOQFWAMFLKKAAYSWREKWAQVVLPLTCVTIALLAINYSSEL 930
QY 325 -DKPLI-----EKLAE-----IYVNSFYKETAELHQ-----351
DB 931 FDDPMLRLTLGEVGRTVVPFSGTSQLGQQLSEHLKDALQAEQEPREVLGDLEFLIF 990
QY 352 ---LSGGEKKKITVPEKISYTTSPCHQLRWVSKRSFKNLLGNPQ-----ASIAQI 399
DB 991 RASVEGGGFNERCLV-----AASF-----RDVGERTVVALFNQAYHSPATALAVDNL 1040
QY 400 IVTVVLGLVIGAIYFGLKNDSTGQ-----NRAGVLPELTNNQCFSS-----SVSAV 445
DB 1041 LFKLCLGPHASIVSNFPQPSALQAKDQFNEGRKGFDIALLNLFFAMAFLASTSFILAV 1100
QY 446 ELFWVEKLLFIH-EYISGYRVSSYPLGKLLSDLLPMRLPSIIIFTCIVYFVMLGLKPRAD 504
DB 1101 SERAVQAK---HVQFVSGVH-VASFWLSALLWDLISF-LIPSULL--LVWF-----K 1145
QY 505 AFVVMFTLMVAYGASSMALAIAAGQSVSVATLMTTCFVFMIFSGLLNLNLTIASW 564
DB 1146 AFDVRAFT--RDGHMADTLLLLLYGWAIIPIMLMN-----FFFLGAATAYTRLTIFNI 1198
QY 565 LSLMLOYF-----SIPRYGF-----ITALQNEFLGQFCPLGNATGNPNVAT---CTGE 611
DB 1199 LSGIATFLMTIMRIPAVKLEELSKTLDHFVLVLPNHLGCM-AVSSFYENYETRYCTSS 1257
QY 612 E-----YLVKQGDIDLSW---GLWKNHVALACMIVIFLTIAYLKLIFL 651
DB 1258 EVAHYCKKYNIOYQENFYANAPGVGRFVVASAAS-----CCAYLILLLFL 1303

RESULT 6
US-08-762-500-75

Sequence 75, Application US/08762500
Patent No. 6030806
GENERAL INFORMATION:
APPLICANT: Landes, Gregory M.
APPLICANT: Burn, Timothy C.
APPLICANT: Connors, Timothy D.
APPLICANT: Dackowski, William R.
APPLICANT: Van Raay, Terence J.
APPLICANT: Klinger, Katherine W.
TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,
METHODS OF MAKING AND USING SAME
TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,
METHODS OF MAKING AND USING SAME
NUMBER OF SEQUENCES: 83
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENZYME CORPORATION
STREET: One Mountain Road
CITY: Framingham
STATE: Massachusetts
COUNTRY: United States of America
ZIP: 01701
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/762,500
FILING DATE: 09-DEC-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/665,259
FILING DATE: 17-JUN-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/10469
FILING DATE: 17-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Dugan, Deborah A.
REGISTRATION NUMBER: 37,315
REFERENCE/DOCKET NUMBER: IGS-9.3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (508) 872-8400
TELEFAX: (508) 872-5415
INFORMATION FOR SEQ ID NO: 75:
SEQUENCE CHARACTERISTICS:
LENGTH: 1704 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-762-500-75

Query Match 7.1%; Score 236.5; DB 3; Length 1704;
Best Local Similarity 20.3%; Pred. No. 9.1e-15;
Matches 169; Conservative 116; Mismatches 241; Indels 305; Gaps 39;

QY 66 NINGIMKPGNLAILGPTGGKSSLDVLAARKDPSGLSGDVLING---APRPANFKNSG 122
DB 533 NLN-LYEGQITVLLGHNGAGKTTTSLMLTGLFPPT--SGRAVISGEYSQDMVQIRKSLG 609
QY 123 YVQDDVVMGTLTVRENLOFSAALRLATTMTNHEKNERINRVIEELGLDKVADSKVGTQF 182
DB 610 LCPQHDILFDNLTVAEHLVYFAQLK---GLSRQKCPPEVKQMLHIIGLEDKWSR-----661
QY 183 IRGVSGGERKRTSIGMELITDPSILSLDEPTTGLSDSTANAVLLLLKRMKSGQRTIIFSI 242
DB 662 SRFLSGMRKLSIGIALIAGSKVLIDLEPTSGMDAISRAIWDLLQR-QKSDRIIVLT 720
QY 243 HQPRYSIFK---LFDSTLLASGLMHPGA---QEALGYFESAGYHC-----BAYNPP 290
DB 721 H-----FMDEADLLGDRIAMWAGELQCCGSSFLKQYK---AGYHMTLVKPEHCNP 770
QY 291 AD-----FFLD-----296
DB 771 EDISQVHHVHPNATLESSAGAEISFILPRESTRHFEGLFAKLEKKQKELGIASFASIT 830

```

297  - - - - - IING 300
Db 831 TBEVFLURVGLVDSSMDIOAIOLPALQYQHERRASDWVDSNLCGAMPDSDGIGALIE 890
Qy 301 DSTAVALN- - - - - REEDFKATEIIEP- - - - - SKQ- 324
Db 891 ERTAVKLTGLALHCOQFWAMFLKKAYSWREKWAQVVLVPLTCTVTALLAINYSSEL 950
Qy 325 -DKPLI- - - - - EKLAE- - - - - IYVNSSFYKTKAELHQ- - - - - 351
Db 951 FDDPMLRLTLGEYGRTVVPFSPGTSQGGQOLSEHLKDALQAESQEBREVLGDLERFLIF 1010
Qy 352 - - - - - LSGEKKKKKTVFKIEISYTSFCHOLRWKSRSPFNLLGNPQ- - - - - ASIAQI 399
Db 1011 RASVEGGGFNERCLV- - - - - AASF- - - - - RDVGERTVVNALFNNQAYHSPATALAVVDNL 1060
Qy 400 IVTWVLGLVIGATYFGLKNDSTGIQ- - - - - NRGVLVFLTTNOCFS- - - - - SVSVA 445
Db 1061 LFKLLCGPHASIVVSNFPQPRSAQAOKDFNEGRKGFDAIMLLFAMAFLASTFFSILAV 1120
Qy 446 ELFWVEKKLFH- EYISGYRVSSYFGLKLLSLLPMRMPLPSIPTCTIVYFMGLKPKAD 504
Db 1121 SERAVQAK- - - - - HQVQSGVH- VASFMLSALLWDLISP- LIPSLLL- - - - - LVVF- - - - - K 1165
Qy 505 AFFVMTFTLMWAYSASSMALATAAGOSVSVATLLMTICFVFMFMIFSGLLLVNLTIASW 564
Db 1166 AFDVRAFT- - - - - RDGHMADTLLLLLYGWAIIFLMYLMN- - - - - PFFLGAATAYTRLTIFNI 1218
Qy 565 LSLMLQVF- - - - - SIPRYGF- - - - - TALQHNEFLGQFCPLGNATGNPCNYAT- - - - - CTGE 611
Db 1219 LSGIATFLWYIMRIPAVKLEELSKLDHVLVLPNHCLGM- AVSSFYENYETRRYCTSS 1277
Qy 612 E- - - - - YLVKQGDLSPW- - - - - GLMKNHVALACMIVIELTAYIKLLEP 651
Db 1278 EVAHYCKKNIQYQENFYAWSAPGVGRFVASMAAS- - - - - GCAYLILFL 1323

RESULT 7
US-09-134-001C-3641
; Sequence 3641, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHY-
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3641
; LENGTH: 273
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3641

```

	Qy	165	I E L G L D K V A D S K V Q T F I R G V S G E R K R T S I G M E L I T D P S I I S L D S P T T G L D S S T A N A V 224 :::
	D b	129	L E L V L G L E E K A S A ----- Y P K O J S G Q K Q V A I A R A L A N E P K V L L C D E A T S A L D P K T T L Q I 183 :::
	Qy	225	L A L L K E M S K Q - C R T I I F S H Q P R V S I F X L P D S L T L A S G R L M F H G P A Q E A L G V F E S A G V H 283 :::
	D b	184	L A L L K K L N K Q L G T I V L I T H E M Q V - V K E I C N K V A M E D G R I V ----- B K G S S 229 :::
	Qy	284	C E A Y A N P ----- A D F F L O I I N G D --- S T A V A L N R E E D F K A T E - I I E P S ---- K O D K P L I T E 330 :::
	D b	230	I Q I P S N P E E L T K D I R T A T H L D Q A L E T I I I A H S A F A Q I A N K W L V E L S Y I G N Q T N E P L I A 289 :::
	Qy	331	K U A E ----- I Y N S S F Y K E T K -- A E L H O L S G G E K K K I T V P K E I S Y 369 :::
	D b	290	H Y S K Y O V T A N I Y G N V E L L O B T P L G S L I V T A G E T K O R K ----- K A L D Y 334 :::

RESULT 9

```

US-09-134-001C-4600
; Sequence 4600, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: LYNN DOUCETTE-STAMM ET AL
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GPC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4500
; LENGTH: 256
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4600

```

Query Match	6.8%;	Score 227;	DB 4;	Length 256;
Best Local Similarity	32.8%;	Pred. No. 3.6e-15;		
Matches 63;	Conservative 40;	Mismatches 69;	Indels 20;	Gaps 7;
Qy	64	LSNIN-CIMKPGLNAILPTGGGKSSLLDLVLRKDPGSLSGDVLINGA-----BRPA	115	
Db	26	LNINISFVEKEGFVAVNGPSGKTYLLNLVLSIDTTISG--GTVVEGKINKLSHKEVA	83	
Qy	116	NP-KCNSGYVVDVVMGTLTVRENLQFSAAALRLATTWTHNKRVERINRVIELGLDKVA	174	
Db	84	NFRKQHLGFTFDYSVLPTTLTVKENIMLPLSVQ---KFKHYEMEQNTKEVAEALGI	136	
Qy	175	DSKVGQTQIRGVSGGERKRTSIGMELTIDPSILSLDEPTTGLOSSTANAVILLIKRMKSQ	234	
Db	137	YNLGNKYPSEISGGQQRATAARAFAVHKPTIIIFADEPTGALDSKSAQDLHLRLDWNKQ	195	
Qy	235	-GRTIIFSIHQP	245	
Db	196	ENSTIMMVTHDP	207	

RESULT 10

```

RESULTS 10
US-09-134-001C-3439
; Sequence 3439, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964

```

```
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3439
; LENGTH: 382
; TYPE: PRMT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3439

      Query Match          6.7%; Score 225.5; DB 4; Length 382;
Best Local Similarity    25.4%; Pred. No. Ie-14;
Matches      87; Conservative 65; Mismatches 122; Indels 69; Gaps 14;

QY   36 VLSPHNICYRVKLKSGFLPCRKPVEIKLSNINIGMKPG-LNAILGPTGGKSLLDVLA 94
     :|||:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB   22 LLSFKDV-----SKGF-----EDVQLNEINIDIEFGFYTLGLPSGCCGTIILKIIA 69

QY   95 ARKDPSGLSGDVLINGAP---RPNPFKNCSGYVVQDDVMVGTLTVRENIQFSAALRLAT 150
     :|||:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB   70 GREYPD--SGDIIVKDKPIGKMPPNKRKVNT--VFQDYALFPHLVNFVDNIAYGLKKLKS 125

QY   151 TWTNHKEKRINRVIELGLDKVADSKVGTOFTRGVSGGERKRTSICMELITPDSIISLD 210
     :|||:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB   126 -----KSEIKRKVTAEALQVLK--SOYEHRQIQCGMSGGQQRVAIRAVNEPEILLDD 177

QY   211 EPTTGDSSPANAVLLLKEM-SKOGRTIFSTHQPRYSIFKLFDSLTLLASGRLMFHGP 269
     :|||:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB   178 ESLSALDLKURTEMQYLLRELQSRGLGITFIFFTHDQEAL-AUSDYIFVMKGDKI----- 231

QY   270 AQEALGVFSAGYHCAYNNPADFFDIINDGSTVALNREEDF----- 313
     :|||:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB   232 -----QQFGTPIDIYDEPNRFVADFGEISNVHGHTWVEDFVNVIYQNFDVCVDNGI 283

QY   314 ----KATEIIEPSQDKPLIEK---LAEIYVNSSFYKETKAEL 349
     :|||:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB   284 KENKKEVETWRP--EDISILVSONDGLFKAKVDSMLFRGVHYEI 324
```

RESULT 11

```

US-08-232-537-2
; Sequence 2, Application US/08232537
; Patent No. 5516655
; GENERAL INFORMATION:
; APPLICANT: Peery, Robert B.
; APPLICANT: Skatrud, Paul L.
; TITLE OF INVENTION: MULTIPLE DRUG RESISTANCE GENE OF
; TITLE OF INVENTION: AUREOBASIDIUM PULLULANS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,537
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X9212
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; TELEFAX: 317-276-1917
; INFORMATION FOR SEQ ID NO: 2:

```

SEQUENCE CHARACTERISTICS:
LENGTH: 1302 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-232-537-2

Query Match 6.7%; Score 224.5; DB 1; Length 1302;
Best Local Similarity 20.4%; Pred. No. 1e-13;
Matches 139; Conservative 102; Mismatches 231; Indels 209; Gaps 27;
QY 4 SNVEVFPVSGQNTGFPATVNDLKAFTGAVLSFHNIICRYVKLSGFLPCRKPVEKEI 63
Db 383 SEINYSAGQBAT-----ESDMK-----ADLVFRNVTF-----VYPAR--TSARA 421
QY 64 LSNINGMKPG-LNAILPTGGGKSLDLVLAARKDPSGLSDVGLNGAPRANPKCNS- 121
Db 422 LEEMSLILKAGOMNAIVTSGCGKTLVSLRLRLDIS--SQLTI-GSHDIDKFNVRSL 478
QY 122 ---GYVQDDVVMGTLTVRENIQFSAALRLATMTNHEKNERINRVIEELGDKV----- 173
Db 479 RRYTALVDQDSVLFSGVLENISYGLG---EHSLSDDVVLERCTEAAKAAALDFVDELPO 535
QY 174 -ADSKVGTQFIRGVSGGKRTSIGMELITDPSILSLDEPTTGLSDSTANAVLLKRM 232
Db 536 GIHTRIGNGYTSLSGGQNRICLAPALVKPALLLDEPTAALDANSEGLIMDAVKVA 595
QY 233 KQGRTIIFSIHQPRYSIFKLPDSLTLLASGRMLFHPGPAQEA---GYFESAGYHCEAYN 288
Db 596 ATGTTVMVAH--RLSTVSDSNIVLMGAGKVIQGNHDELMOLEGAYFNL--IOAQQLN 651
QY 289 NPADFLLDIINGDSTAVA---LNREDFKA---TBIEEP--SKQDKPLIEKLAETVYVNSS 340
Db 652 DADSSAESAATTSQVTPQKASKSDSAASDSTETVPPQAKKEDKPA----- 699
QY 341 FYKETAEHLQSGCKKKITVFEISYITTSFCHQLRWVSKRSPKLLGNPQASIAQII 400
Db 700 -----KKAGFWKLLRLCLAKSDSPIIALG-----LAASI 730
QY 401 VTVVLGLVIG-AIYFGLKNDSTGIONRAGVLPFLTNQCFSSVAVELFVVEKKLFIHEY 459
Db 731 VSG--GIILGEAIVFG-----NLISVNDLESFD----- 757
QY 460 ISGYRVSSYFLGKLLSLLPMRLPSIIFTCIVYFMLGLKPKADAFFVMPFTLMVA-- 517
Db 758 -----FRSRADLFSLFFILALIF 778
QY 518 -YSASSMALATAGOSVSVATL-LMTICFVPMIFSG-----LLVNLTTIASWLSWLO 570
Db 779 SYAGNCCFCGIVSSHFAKIQHISLASILRQDMQWFGSQSVPSLSSDAGQLACLSG 838
QY 571 FSIIPRYGPTALQHNFEFLGNFCPGLNATGNPCNYATCTGEBYLVKQGLDLSWGLK 630
Db 839 VAI-----GTIFVCVSIIG-----GIILAHVVAWKIA 866
QY 631 VALACMIVIFLTAYLKLFL 651
Db 867 VVLLAAVPMITAGYVRLVL 887

RESULT 12
US-09-107-532A-3849
; Sequence 3849, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham

STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 3849:
SEQUENCE CHARACTERISTICS:
LENGTH: 345 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...345
SEQUENCE DESCRIPTION: SEQ ID NO: 3849:
US-09-107-532A-3849

Query Match 6.6%; Score 222.5; DB 4; Length 345;
Best Local Similarity 27.5%; Pred. No. 1.8e-14;
Matches 96; Conservative 58; Mismatches 124; Indels 71; Gaps 18;
QY 60 EKEILSNINGIM--KPGLNAILGPTGGKSSLLDLVLAARKDPSGLSDVLI---NGAPRP 114
Db 16 EKQILA-IDEQIEKGELVTLGFGGCKSLRCITGLEQPK--SGEIVLDNENIADKP 72
QY 115 ANFKCNGYVQDDVVMGTLTVRENIQFSAALRLATMTNHEKNERINRVIEELGL-DKV 173
Db 73 TKDR-NIGFVFPQYALPPTMTVFENVAFGLKVKLSPEVIQEK-----VFEMLSLVDMT 125
QY 174 ADSKVGTFIRGVSGGKRTSIGMELITDPSILSLDEPTTGLSDSTANAVLLKRM 233
Db 126 EQADKNVQFL---SGGQRVALARSVTEPKVLLDDEPLSALDARIRKQIQORDLRAIQ 182
QY 234 Q-GRITIFSIHQPRYSIFKLPDSLTLLASGRMLFHPGPAQEAALGYFESAGYHCEAYN 292
Db 183 SLGWTMIFVTHDQBEAM-RISDRIFVMEAGRV-----AQVSTSK-----ELYQNQS 228
QY 293 FFLDIINGDSTAVALENREEDFKATEII-EPKQDKPLIEKLAETVYVNSSPKYKAEHLQ 351
Db 229 RFVAEFTGN-----YNRPFWYELNQYTHFSAKNRCIYYL-----RPELIQ 270
QY 352 LSGGKCKKITVFEISYITTSFCHQLRWVSKRSPKLLGNPQASIAQII 400
Db 271 FEPVQGVKIPVL-----W--KESF--ILGNIQRYVFTQL 301

RESULT 13
US-08-612-734B-2
; Sequence 2, Application US/08612734B
; Patent No. 5914246
; GENERAL INFORMATION:
; APPLICANT: Peery, Robert B.

APPLICANT: Skatrud, Paul L.
APPLICANT: Tobin, Matthew B.
TITLE OF INVENTION: Multiple Drug Resistance Gene of
TITLE OF INVENTION: Aspergillus fumigatus
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Eli Lilly and Company
STREET: Lilly Corporate Center, DC1501
CITY: Indianapolis
STATE: Indiana
COUNTRY: U.S.A.
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/612,734B
FILING DATE: 08-MAR-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Craig, Anne I.
REGISTRATION NUMBER: 32,976
REFERENCE/DOCKET NUMBER: X-9681
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-354-9570
TELEFAX: 617-354-4043
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1349 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-612-734B-2

Query Match 6.5%; Score 219; DB 2; Length 1349;
Best Local Similarity 21.9%; Pred. No. 4.2e-13;
Matches 154; Conservative 105; Mismatches 257; Indels 188; Gaps 34;

QY 11 PVSQNTNGFPATVS-----NDLKFT-EGAVL-----SFHNICRYVKLSGFLP 54
DB 395 PNGAFTNGVAAAKIYTIIDRRSPLDYSDGKVLDPHEGNIERNVXH-----IYP 447
QY 55 CRKPEKEILSNINGMKPG-LNAILGPTGGKSSLLDVLARKPSGLSGDLVINGAP- 112
DB 448 SRP--EVTVMEDVLSMPAGKTTALVPGSGSKSTVVGVERFYLPVG--GQVLLDGHDI 503
QY 113 RPAFKC---NSGYVVQDDVVMGTLTVRENLFSAALRLATTMTNHEKNERINRVIEE-- 167
DB 504 QTLNLRLRQQLSLYSQEPVLEST-TIFRNIHGH-----LIGTKFHESKDKIRLVENNA 558
QY 168 -----LGLDKVADSKVGTQFIRG--VSGGRKRTSIGMELITDPSILSLDEPTTGL 216
DB 559 RMANAHDFTMALPEGYDVTNVGQ--RGFLSGGQKQRIARAIVSDPKILLDEATSAI 615
QY 217 DSTANAVLLLLKRSKQRTIIFSTHQPRYSIFKPLDSLTLASRLMFMHPQAEALG- 275
DB 616 DTKSEGVOALDK-AAEGRTTIVIAH--RLSTIKTAHNVAMVGKTAEOQTHDELVD 672
QY 276 ---YF-----ESAGYHCAYNNPAPDFFLDINGDSTAVALNREED----- 312
DB 673 KGYTKLVLEAQRINEKEAEALDMDADDPGQGVTRIKTAVSSNSLDAVDEKARL 732
QY 313 -PKATEIIPSKODKPLIEKLAIEYVNSSFYKTK-----AEHLQL-----SG 354
DB 733 EMKRTG-TQKSVSSAVLSKKVPEQEKSLTLVLPFGAFNRPGLGYMLIGLTFSLAGG 791
QY 355 GEKKKKITVFEISYTT-----SFCHOLRWVSKSEFNKLLGNPOASTAQIIVTVVLGIVCA 411
DB 792 GQPTQAFLYAKAISTLSLPESMFHKLRL--HDANFWSLMFF-VVGIAQFISLSINGTAFAI 848
QY 412 I-----YFGLKNDSTGIONRAGVLPFLTTNQCFSVSVAVELFV 449

Db 849.CSERLIRARSOAFRSILRQDISPFDREENGSTG---ALTSLSTETTKNLSGVSGVTL-- 902
QY 450 VEKKLFPIHEYISGYRVSSYFGLKLLSLLPMRMPLPSIIFTCIVYFMGLKPKADAF--F 507
DB 903 -----GTIIMTSTTGAAMIITAIAGWKALVCISVVPILL-----ACGFLRF 945
QY 508 VMMPTLMVAVSA--SSMALAIAGQSVSVATL-----LMTIC 544
DB 946 YMLAQFOQRSKSAVEGSASYACEATSAIRTVASLTREODVGVYHDDQLQKGRKSLISVL 1005
QY 545 FVFMWIFS--GLLVNLTTIASWLSWLYFSIPRYGFTALQHNFF 586
DB 1006 RSSLLYASSQALVFCVALGFW-----YGGTLIGHEY 1038

RESULT 14
US-09-107-532A-4844
Sequence 4844, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 4844:
SEQUENCE CHARACTERISTICS:
LENGTH: 329 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...329
SEQUENCE DESCRIPTION: SEQ ID NO: 4844:
US-09-107-532A-4844

Query Match 6.4%; Score 216; DB 4; Length 329;
Best Local Similarity 24.2%; Pred. No. 7.9e-14;
Matches 80; Conservative 58; Mismatches 129; Indels 64; Gaps 12;
QY 61 KEILSNINGIMKPG-LNAILGPTGGKSSLLDVLARKDPSGLSGDLVINGAP-----RPA 115

Search completed: August 1, 2003, 18:52:39
Job time : 39.485 secs

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.
OM protein - protein search, using sw model
Run on: August 1, 2003, 18:38:41 ; Search time 5.49875 Seconds
(without alignments)
1257.181 Million cell updates/sec
Title: US-09-856-927-4
Perfect score: 769
Sequence: 1 FGLGAEYATSSMALAIATG.....MIIIFLTATYKLFLFKYS 147
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 127863 seqs, 47026705 residues
Total number of hits satisfying chosen parameters: 127863
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	623	81.0	655	1 ABG2 HUMAN	Q9unq0 homo sapien
2	414.5	53.9	650	1 ABG3 MOUSE	Q99p81 mus musculus
3	128	16.6	687	1 WHIT_DROME	P10090 drosophila
4	124.5	16.2	1499	1 CDR2 CANAL	P78595 candida alb
5	123.5	16.1	646	1 ABG4 HUMAN	Q9h172 homo sapien
6	113	14.7	678	1 ABG1 HUMAN	P45844 homo sapien
7	110.5	14.4	1333	1 YN99 YEAST	P53756 saccharomyc
8	110	14.3	666	1 ABG1 MOUSE	Q64343 mus musculus
9	109.5	14.2	1501	1 CDR1 CANAL	P43071 candida alb
10	109.5	14.2	1501	1 SNQ2 YEAST	P32568 saccharomyc
11	108.5	14.1	598	1 YPC3 CAEEL	Q11180 caenorhabdi
12	108	14.0	1294	1 YOH5 YEAST	Q08234 saccharomyc
13	105	13.7	679	1 WHIT_CERCA	Q17320 ceratitidis c
14	102.5	13.3	1564	1 PDBA YEAST	P51533 saccharomyc
15	99.5	12.9	1410	1 PDRB YEAST	P40550 saccharomyc
16	97	12.6	1501	1 CDR3 CANAL	Q042690 candida alb
17	94.5	12.3	610	1 YQSC CAEEL	Q05360 lucilia cup
18	94.5	12.3	677	1 WHIT_LUCCU	Q09466 caenorhabdi
19	91.5	11.9	1530	1 BPR1 SCHPO	P41820 schizosacch
20	87	11.3	1529	1 PORF YEAST	Q04182 saccharomyc
21	86	11.2	1490	1 CDR4 CANAL	Q74676 candida alb
22	85	11.1	695	1 WHIT ANOAG	Q27256 anopheles g
23	83.5	10.9	1511	1 PDR5 YEAST	P33302 saccharomyc
24	82.5	10.7	651	1 ABG5 HUMAN	Q9b222 homo sapien
25	81.5	10.6	652	1 ABG5 MOUSE	Q99p81 mus musculus
26	81.5	10.6	1511	1 PDR3 YEAST	Q02785 saccharomyc
27	81	10.5	672	1 ABG8 RAT	P58428 rattus norv
28	80	10.4	561	1 B1O5 YEAST	P53744 saccharomyc
29	79.5	10.3	668	1 BROW DROVI	Q24739 drosophila
30	79	10.3	-275	1 YD84 YEAST	Q12359 saccharomyc
31	79	10.3	839	1 NAH3 DIDMA	Q28362 didelphis m
32	78.5	9.9	709	1 WHIT ANOAL	Q16928 anopheles a
33	76.5	9.9	364	1 YHIM_ECOLI	P37630 escherichia

34	76.5	9.9	652	1 ABG5 RAT	Q99pe7 rattus norv
35	76.5	9.9	656	1 YID7 YEAST	P40534 saccharomyc
36	75.5	9.8	518	1 SP5B BACSU	Q00758 bacillus su
37	74.5	9.7	142	1 YNKS YEAST	P50943 saccharomyc
38	74.5	9.7	1049	1 ADP1 YEAST	P25371 saccharomyc
39	74	9.6	149	1 DISD DICDI	P02888 dictyosteli
40	74	9.6	253	1 DISC DICDI	P02887 dictyosteli
41	74	9.6	482	1 YF1G BACSU	P54723 bacillus su
42	74	9.6	831	1 NAH3 RAT	P26433 rattus norv
43	73	9.5	527	1 COX1 ARATH	Q07063 arabidopsis
44	73	9.5	590	1 CAN1 YEAST	P04817 saccharomyc
45	73	9.5	673	1 ABG8 MOUSE	Q9dbm0 mus musculu

ALIGNMENTS

RESULT 1

ABG2 HUMAN

ID ABG2 HUMAN STANDARD; .PRT; 655 AA.

AC Q9UNQ0; Q95374; Q9BY73; Q9NUS0;

DT 16-OCT-2001 (Rel. 40, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE ATP-binding cassette, sub-family G, member 2 (Placenta-specific ATP-binding cassette transporter) (Breast cancer resistance protein).

GN ABCG2 OR ABCP OR BCRP OR BCRP1.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RP SEQUENCE FROM N.A.

RC TISSUE=Placenta;

RX MEDLINE=95065313; PubMed=9850061;

RA Allikets R., Schriml L.M., Hutchinson A., Romano-Spica V., Dean M.;

RT "A human placenta-specific ATP-binding cassette gene (ABCP) on chromosome 4q22 that is involved in multidrug resistance.";

RL Cancer Res. 58:5337-5339(1998).

[2]

RP SEQUENCE FROM N.A.

RC TISSUE=Breast cancer;

RX MEDLINE=99080071; PubMed=9861027;

RA Doyle L.A., Yang W., Abruzzo L.V., Krogmann T., Gao Y., Rishi A.K.,

RT Ross D.D.;

RT "A multidrug resistance transporter from human MCF-7 breast cancer cells.";

RL Proc. Natl. Acad. Sci. U.S.A. 95:15665-15670(1998).

[3]

RP ERRATUM.

RA Doyle L.A., Yang W., Abruzzo L.V., Krogmann T., Gao Y., Rishi A.K.,

RT Ross D.D.;

RL Proc. Natl. Acad. Sci. U.S.A. 96:2569-2569(1999).

[4]

RP SEQUENCE FROM N.A.

RA Kage K., Tsukahara S., Sugiyama T., Asada S., Ishikawa E., Tsuruo T.,

RT Sugimoto Y.;

RT "Breast cancer resistance protein constitutes a 140-kDa complex as a homodimer.";

RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.

[5]

RP SEQUENCE OF 198-655 FROM N.A.

RC TISSUE=Placenta;

RA Isogai T., Oka T., Hayashi K., Sugiyama T., Otauki T., Suzuki Y.,

RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H., Sugawara M.,

RA Wagatsuma M., Hooiriri T., Kaku Y., Kodaira H., Kondo H., Takiguchi S.,

RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,

RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,

RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,

RA Ninomiya K., Iwavanagi T.;

RT "NEDO human cDNA sequencing project.";

RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.

[6]

REVIEW.
 RX MEDLINE=21474438; PubMed=11590207;
 RA Schmitz G., Langmann T., Heimel S.;
 RT "Role of ABCG1 and other ABCG family members in lipid metabolism."; J. Lipid Res. 42:1513-1520(2001).
 RL -!- FUNCTION: XENOBOTIC TRANSPORTER THAT APPEARS TO PLAY A MAJOR ROLE IN THE MULTIDRUG RESISTANCE PHENOTYPE OF A SPECIFIC MCF-7 BREAST CANCER CELL LINE. WHEN OVEREXPRESSED, THE TRANSPORTED CELLS BECOME RESISTANT TO MITOXANTHONE, DAUNORUBICIN AND DOXORUBICIN, DISPLAY DIMINISHED INTRACELLULAR ACCUMULATION OF DAUNORUBICIN, AND MANIFEST AN ATP-DEPENDENT INCREASE IN THE EFFLUX OF RHODAMINE 123.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCG (WHITE) SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF103796; AAD09188.1; -;
 DR EMBL; AF098951; AAC97367.1; -;
 DR EMBL; AB056867; BAB39212.1; -;
 DR EMBL; AK002040; BAA92050.1; -;
 DR Genew; HGNC:74; ABCG2.
 DR MIM; 603756; -;
 DR GO; GO:0016021; C:integral to membrane; TAS.
 DR GO; GO:0005524; F:ATP-binding activity; TAS.
 DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . . ; TAS.
 DR GO; GO:0005215; F:transporter activity; TAS.
 DR GO; GO:0008559; F:xenobiotic-transporting ATPase activity; TAS.
 DR GO; GO:0009315; P:drug resistance; TAS.
 DR GO; GO:0006832; P:small molecule transport; TAS.
 DR InterPro; IPR003593; AAA ATPase.
 DR InterPro; IPR003439; ABC transporter.
 DR Pfam; PF00005; ABC_tran; 1
 DR ProDom; PD000006; ABC_transporter; 1.
 DR SMART; SM00382; AAA; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER_1; FALSE_NEG.
 DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
 KW ATP-binding; Transmembrane; Transport.
 FT DOMAIN 1 395 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 396 416 POTENTIAL.
 FT DOMAIN 417 428 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 429 449 POTENTIAL.
 FT DOMAIN 450 477 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 478 498 POTENTIAL.
 FT DOMAIN 499 506 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 507 527 POTENTIAL.
 FT DOMAIN 528 535 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 536 556 POTENTIAL.
 FT DOMAIN 557 630 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 631 651 POTENTIAL.
 FT DOMAIN 652 655 CYTOPLASMIC (POTENTIAL).
 FT NP_BIND 80 87 ATP (POTENTIAL).
 FT CARBOHYD 418 418 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 557 557 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 596 596 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 24 24 V -> A (IN REF. 2 AND 4).
 FT CONFLICT 166 166 F -> Q (IN REF. 2 AND 4).
 FT CONFLICT 208 208 F -> S (IN REF. 1).
 FT CONFLICT 315 316 MISSING (IN REF. 5).
 FT CONFLICT 482 482 R -> T (IN REF. 2).
 SQ SEQUENCE 655 AA; 72343 MW; 89A6D3511DC5CCE0 CRC64;
 Query Match 81.0%; Score 623; DB 1; Length 655;
 Best Local Similarity 81.6%; Pred. No. 2.9e-48;
 Matches 120; Conservative 8; Mismatches 17; Indels 2; Gaps 1;
 QY 1 FGLGAEAYTASSMALAIATGQSVSVATLLMTIAFVFMFLFSGLLVNLRTTIGPMLSWLQY 60

Db 511 FTLMWVAYSASSMALAIATGQSVSVATLLMTIAFVFMFLFSGLLVNLRTTIGPMLSWLQY 570
 QY 61 FSPIRYGTALQYNEFLGQEFPCPGNVTNDSTCNVSYAICTGNEYLNQGLSPLWGLWK 120
 Db 571 FSPIRYGTALQYNEFLGQEFPCPGNVTNDSTCNVSYAICTGNEYLNQGLSPLWGLWK 628
 QY 121 NHVALACMIIFLTITAYLKLLFLKKYS 147
 Db 629 NHVALACMIIFLTITAYLKLLFLKKYS 655
 RESULT 2
 ABC3 MOUSE
 ID ABC3_MOUSE STANDARD; PRT; 650 AA.
 AC Q99P81;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE ATP-binding cassette, sub-family G, member 3.
 GN ABCG3.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Spleen;
 RA MEDLINE=21030753; PubMed=11178751;
 RX Mickley L., Jain P., Miyake K., Schriml L.M., Rao K., Fojo T.,
 RA Bates S., Dean M.;
 RT "An ATP-binding cassette gene (ABCG3) closely related to the multidrug transporter ABCG2 (MXR/ABCP) has an unusual ATP-binding domain.";
 RT Mamm. Genome 12:86-88(2001).
 RL -!- SUBUNIT: May dimerize with another subunit to form a functional transporter.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -!- TISSUE SPECIFICITY: Highest levels of expression in thymus and spleen. Detected in lung and small intestine.
 CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCG (WHITE) SUBFAMILY.
 CC -!- CAUTION: Seems to have a defective ATP-binding region.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF324242; AAK14241.1; -;
 DR MGD; MGI:1351624; Abcg3.
 DR InterPro; IPR003439; ABC transporter.
 DR Pfam; PF00005; ABC_tran; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER_1; FALSE_NEG.
 DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
 KW Transmembrane; transport.
 FT DOMAIN 1 387 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 388 408 1 (POTENTIAL).
 FT DOMAIN 409 420 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 421 441 2 (POTENTIAL).
 FT DOMAIN 442 469 3 (POTENTIAL).
 FT TRANSMEM 470 490 4 (POTENTIAL).
 FT DOMAIN 491 498 5 (POTENTIAL).
 FT TRANSMEM 499 519 6 (POTENTIAL).
 FT DOMAIN 520 527 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 528 548 5 (POTENTIAL).
 FT DOMAIN 549 623 6 (POTENTIAL).
 FT TRANSMEM 624 644 6 (POTENTIAL).
 FT DOMAIN 645 648 CYTOPLASMIC (POTENTIAL).
 SQ SEQUENCE 650 AA; 73623 MW; 86A5ABB4DD26971C CRC64;

Query Match 53.9%; Score 414.5; DB 1; Length 650;
 Best Local Similarity 58.0%; Pred. No. 1.1e-29;
 Matches 80; Conservative 22; Mismatches 35; Indels 1; Gaps 1;

QY 7 AYTSAMALAIATGQSVSVATLLMTIAVFMVLESLGLVNLRTGTGPMVLSWLYQYSIPRY 66
 DB 509 AYASASLPUSIGAGENAVPILLVTIYFVFMFLFSGLSYPGSFLPKLSWLYQYSIPHY 568
 QY 67 GFTALQYNEFLGQFCFGFNVDNSTCVNSYAICTGNEYLINQGLSIPWGLWKNHVALA 126
 DB 569 GFRALLHNEFLQNFQCPHNTSEVRCHN-VYICTGEEFLMQGIDLSWGFWENHLALV 627
 QY 127 CMIIPLTITAYLKLFLK 144
 DB 628 CTMIILLTITVYQLQVK 645

RESULT 3

WHIT_DROME
 ID WHIT DROME STANDARD; PRT; 687 AA.
 AC P10050; Q9V3A2; Q9XY33;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE White protein.
 GN W OR EG:BACN33B1.1 OR CG2759.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxId=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Head;
 RX MEDLINE=90221897; PubMed=2109311;
 RA Pepling M., Mount S.M.;
 RT "Sequence of a cDNA from the Drosophila melanogaster white gene.";
 RL Nucleic Acids Res. 18:1633-1633(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85134865; PubMed=6084717;
 RA O'Hare K., Murphy C., Lewis R., Rubin G.M.;
 RT "DNA sequence of the white locus of Drosophila melanogaster.";
 RL J. Mol. Biol. 180:437-455(1984).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21100348; PubMed=11156992;
 RA Lukatsovich T., Asztalos Z., Awano W., Baba K., Kondo S., Niwa S., Yamamoto D.;
 RT "Dual-tagging gene trap of novel genes in Drosophila melanogaster.";
 RL Genetics 157:727-742(2001).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D., Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F., George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N., Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X., Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D., Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G., Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D., Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M., Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S., Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P., Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I., Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P., de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M., Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P., Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W., Foster C., Gabrielson A.B., Garg N.S., Gelbart W.M., Glasser K., Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C., Jaitali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A., Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z., Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X., Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D., Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A., Mout S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L., Nelson D.L., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M., Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G., Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H., Shue B.C., Siden-Kimios I., Simpson M., Skupski M.P., Smith T., Spier E., Spradling A.C., Stapleton M., Strong R., Sun E., Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X., Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J., Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L., Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O., Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Oregon-R;
 RX MEDLINE=20196011; PubMed=10731137;
 RA Benos P.V., Gatt M.K., Ashburner M., Murphy L., Harris D., Barrell B.G., Ferraz C., Vidal S., Brun C., Demailles J., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S., Galibert F., Borkova D., Minana B., Kafatos F.C., Louis C., Siden-Klamos I., Bolshakov S., Papagiannakis G., Spanos L., Cox S., Madueno E., de Pablo F., Modolell J., Peter A., Schoettler P., Werner M., Mourikios B., Beinert N., Dowe G., Schaefer U., Jaekle H., Bucheton A., Callister D.M., Campbell L.A., Darlamitsou A., Henderson N.S., McMillan P.J., Salles C., Tait E.A., Valentini P., Saunders R.D.C., Glover D.M.;
 RA "From sequence to chromosome: the tip of the X chromosome of D. melanogaster.";
 RT Science 287:2220-2222(2000).
 RN [6]
 RP SEQUENCE OF 224-331 FROM N.A.
 RX MEDLINE=8939145; PubMed=2503416;
 RA Tearle R.G., Belote J.M., McKeown M., Baker B.S., Howells A.J.;
 RT "Cloning and characterization of the scarlet gene of Drosophila melanogaster.";
 RL Genetics 122:595-606(1989).
 CC -!- FUNCTION: PART OF A MEMBRANE-SPANNING PERMEASE SYSTEM NECESSARY FOR THE TRANSPORT OF PIGMENT PRECURSORS INTO PIGMENT CELLS
 CC RESPONSIBLE FOR EYE COLOR. WHITE DIMERIZE WITH BROWN FOR THE TRANSPORT OF GUANINE AND WITH SCARLET FOR THE TRANSPORT OF TRYPTOPHAN.
 CC -!- SUBUNIT: HETERODIMER OF WHITE WITH EITHER BROWN OR SCARLET.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC EMBL; X51749; CAA36038.1; -
 DR EMBL; X02974; CAA26716.1; -
 DR EMBL; AB028139; BAA78210.1; -
 DR EMBL; AE003425; AAF45826.1; -
 DR EMBL; AL133506; CAB65847.1; -
 DR EMBL; X76202; CAA53795.1; -
 DR PIR; S08635; FYFFW.
 DR FlyBase; FBgn003996; w.
 DR GO; GO:0004888; P:transmembrane receptor activity; NAS.
 DR GO; GO:0006727; P:omochrome biosynthesis; IMP.
 DR InterPro; IPR003593; AAA ATPase.
 DR InterPro; IPR003439; ABC_transporter.

```

DR InterPro; IPR005284; Pigment_permease.
DR Pfam; PF000005; ABC tran; 1.
DR ProDom; PD000006; ABC transporter; 1.
DR SMART; SM00382; AAA; 1.
DR TIGRfam; TIGR00955; 3a01204; 1.
DR PROSITE; PS00211; ABC TRANSPORTER 1; 1.
DR PROSITE; PS00893; ABC TRANSPORTER 2; 1.
KW Pigment; ATP-binding; Transmembrane; Transport.
FT NP_BIND 130 137 ATP (BY SIMILARITY).
FT TRANSMEM 435 433 POTENTIAL.
FT TRANSMEM 465 485 POTENTIAL.
FT TRANSMEM 515 533 POTENTIAL.
FT TRANSMEM 542 563 POTENTIAL.
FT TRANSMEM 576 594 POTENTIAL.
FT TRANSMEM 659 678 POTENTIAL.
FT CONFLICT 25 29 GDSGA -> LIFEIPHCVTAD (IN REF. 2 AND 3).
FT CONFLICT 49 49 L -> R (IN REF. 4 AND 5).
FT CONFLICT 335 371 VQAQPTNYPADFYVQVLAVVGGREISRRDRIAKIC ->
ITLHNSYPAPVPSVLPTTIRFTTVRCWPLCPDGRSSPVI
GSPPYG (IN REF. 3).
SQ SEQUENCE 687 AA; 75672 MW; 24FAD799DE0D396 CRC64;

Query Match 16.6%; Score 128; DB 1; Length 687;
Best Local Similarity 27.5%; Pred. No. 0.00035;
Matches 36; Conservative 25; Mismatches 52; Indels 18; Gaps 3;

QY 10 ASSMALATATGOSVSVATLLMTIAFVFMFLPSGLLVNLRITGPMWSLQYFSIPRYGFT 69
DB 567 SSSSTSMALSGVPPVI-----IPFLFGFFLNSGVSVPVYLKMLSYLSWFRYANE 615

QY 70 ALQYNEFLGQEP-C-VDNSTCVN-----PGFN-VTDNSTCVN-----SYAICTGNEYLNQIGIELSP 115
DB 616 GLLINQWADVE--PG-----EISCTSSNTTCSSKVKVILETNFSAADLPIDVGLAILI 668

QY 130 IIFLTIAVYKL 140
DB 669 VSRFLAYLAL 679

RESULT 4
CDR2_CANAL STANDARD; PRT; 1499 AA.
AC P78595;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Multidrug resistance protein CDR2.
GN CDR2.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5476;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SC5314;
RX MEDLINE=97195787; PubMed=9043118;
RA Sanglard D., Ischer F., Monod M., Bille J.;
RT "Cloning of Candida albicans genes conferring resistance to azole
antifungal agents: characterization of CDR2, a new multidrug ABC
transporter gene."
RL Microbiology 143:405-416(1997).
RC -!- FUNCTION: MULTIDRUG EFFLUX TRANSPORTER. CONFERS RESISTANCE TO
AZOLE ANTIFUNGAL AGENTS, TO OTHER ANTIFUNGALS (TERBINAFINE,
AMOROLFIN) AND TO A VARIETY OF METABOLIC INHIBITORS.
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. PDR5 SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/

```

```

CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U63812; AAB96797.1; -.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC transporter.
DR InterPro; IPR005285; PDR.
DR Pfam; PF00005; ABC tran; 2.
DR ProDom; PD000006; ABC transporter; 2.
DR SMART; SM00382; AAA; 2.
DR TIGRfam; TIGR00956; 3a01205; 1.
DR PROSITE; PS00211; ABC TRANSPORTER 1; 1.
DR PROSITE; PS00893; ABC TRANSPORTER 2; 2.
KW ATP-binding; Transmembrane; Glycoprotein; Transport; Repeat.
FT DOMAIN 1 511 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 512 532 POTENTIAL.
FT TRANSMEM 546 566 POTENTIAL.
FT TRANSMEM 596 616 POTENTIAL.
FT TRANSMEM 621 641 POTENTIAL.
FT TRANSMEM 660 680 POTENTIAL.
FT TRANSMEM 763 783 POTENTIAL.
FT DOMAIN 784 1193 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 1194 1214 POTENTIAL.
FT TRANSMEM 1229 1249 POTENTIAL.
FT TRANSMEM 1279 1299 POTENTIAL.
FT TRANSMEM 1315 1335 POTENTIAL.
FT TRANSMEM 1354 1374 POTENTIAL.
FT TRANSMEM 1465 1485 POTENTIAL.
FT NP_BIND 893 900 ATP (POTENTIAL).
SQ SEQUENCE 1499 AA; 168892 MW; 425562D690AD6A9F CRC64;

Query Match 16.2%; Score 124.5; DB 1; Length 1499;
Best Local Similarity 25.2%; Pred. No. 0.0015;
Matches 41; Conservative 25; Mismatches 66; Indels 31; Gaps 7;

QY 11 SSMLATATGOSVSVATLLMTIAFVFMFLPSGLLVNLRITGPMWSLQYFSIPRYGFTA 70
DB 638 SHMFRSIGAVTTTATAMSLSTVFLLAMIIYAGFVLPPIYILGWSRWIRINPVTYIFES 697

QY 71 LQYNEFLGQEP-C-VDNSTCVN-----PGFN-VTDNSTCVN-----SYAICTGNEYLNQIGIELSP 115
DB 698 LMVNEFHGREFECCQYIPSGGFENLPVENKVCVTGTPGTVVQGTGY-IKLAQYPS 756

QY 116 WGLWKNHVALACMIIFLTIAVYKL-----LFLK 144
DB 757 SHKWNFGITVAFVAVFFLGV-YVALTEKNGASKQGEIVLFLK 798

RESULT 5
ABG4_HUMAN STANDARD; PRT; 646 AA.
AC Q9H172;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE ATP-binding cassette, sub-family G, member 4.
GN ABCG4 OR WHITE2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21515231; PubMed=11606068;
RA Engel T., Lorkowski S., Lueken A., Rust S., Schluter B., Berger G.,
RA Cullen P., Asmann G.;
RT "The human ABCG4 gene is regulated by oxysterols and retinoids in
monocyte-derived macrophages."
RL Biochem. Biophys. Res. Commun. 288:483-488(2001).
RN [2]
RP SEQUENCE OF 20-646 FROM N.A.
RC TISSUE=Dorsal root ganglion;
RX MEDLINE=22170423; PubMed=12183068;
RA Oldfield S., Lowry C., Ruddick J., Lightman S.;

```

RT "ABC4: a novel human white family ABC-transporter expressed in the
 RL brain and eye."
 CC Biochim. Biophys. Acta 1591:175-179 (2002).
 CC -1- FUNCTION: May be involved in macrophage lipid homeostasis.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCG (WHITE)
 CC SUBFAMILY.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

CC EMBL: AJ308237; CAC87131.1; --
 DR EMBL: AJ300465; CAC17140.1; --
 DR PIR: JC7777; JC7777.
 DR Genbank: HGNC:13884; ABCG4.
 DR InterPro: IPR003593; AAA ATPase.
 DR InterPro: IPR003439; ABC transporter.
 DR Pfam: PF00005; ABC trans; 1.
 DR ProDom: PD000006; ABC transporter; 1.
 DR SMART: SM00382; AAA; 1.
 DR PROSITE: PS00211; ABC TRANSPORTER 1; 1.
 DR PROSITE: PS00893; ABC TRANSPORTER 2; 1.
 KW ATP-binding; Glycoprotein; Transmembrane; Transport.
 FT DOMAIN 1 393
 FT TRANSMEM 394 414
 FT DOMAIN 415 425
 FT TRANSMEM 426 446
 FT DOMAIN 447 472
 FT TRANSMEM 473 493
 FT DOMAIN 494 503
 FT TRANSMEM 504 524
 FT DOMAIN 525 532
 FT TRANSMEM 533 553
 FT DOMAIN 554 617
 FT TRANSMEM 618 638
 FT DOMAIN 639 646
 FT NP BIND 102 109
 FT CARBOHYD 422 422
 FT SEQUENCE 646 AA; 71895 MW; 9CCE6E150772611 CRC64;

Query Match
 Best Local Similarity 16.1%; Score 123.5; DB 1; Length 646;
 Matches 38; Conservative 25; Mismatches 68; Indels 11; Gaps 4;
 QY 5 ABAYTASSMALAIATGQSVSVATLMTIAFVFMFLFGLLVNLRITGFWLSWLYQFSIP 64
 Db 512 ATALVAQSLGLLGAASNSLQVATFVGPVTAIPVLLFSGFFVSEKTIPTYLQWSSVLSV 571
 QY 65 RYGFALQVNEFLGQFCFQFNVTNSTCVSYAICTGNE-VLINOGLSLSPWGLWKNHV 123
 Db 572 RYGFEGVILTIY-GME-----RGDLTCLBER--CPFREPOSILRALVDKALYMDPL 621
 QY 124 ALACMIIFLTAYLKLFLKK 145
 Db 622 VLGIFFALRLAYLVLRVK 643

RESULT 6
 ID ABG1 HUMAN STANDARD; PRT: 678 AA.
 AC P45844; Q9BXK6; Q9BXK7; Q9BXK8; Q9BXK9; Q9BXL0; Q9BXL1; Q9BXL2;
 CC Q9BXL3; Q9BXL4;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE ATP-binding cassette, sub-family G, member 1 (White protein homolog)
 DE (ATP-binding cassette transporter 8).
 GN ABCG1 OR ABC8 OR WH1.

OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1] SEQUENCE OF 3-678 FROM N.A. (ISOFORMS 1 AND 4).
 RP TISSUE=Retina;
 RC MEDLINE=96256850; PubMed=8659545;
 RA Chen H.M., Rossier C., Lalliot M.D., Lynn A., Chakravarti A.,
 RA Perrin G., Antonarakis S.E.;
 RT "Cloning of the cDNA for a human homologue of the Drosophila white
 RL gene and mapping to chromosome 21q22.3.";
 RN Am. J. Hum. Genet. 59:66-75 (1996).
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RX MEDLINE=2028799; PubMed=10830953;
 RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
 RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Groner Y.,
 RA Soeda E., Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K.,
 RA Polley A., Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
 RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
 RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
 RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
 RA Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
 RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecker H.,
 RA Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,
 RA Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis P.,
 RA Lehrach H., Reinhardt R., Yaspo M.-L.;
 RT "The DNA sequence of human chromosome 21.";
 RL Nature 405:311-319 (2000).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RX MEDLINE=20408883; PubMed=10950923;
 RA Berry A., Scott H.S., Kudoh J., Tallor I., Korostishevsky M.,
 RA Wattenhofer M., Guipponi M., Barras C., Rossier C., Shibuya K.,
 RA Wang J., Kawasaki S.E., Bonne-Tamir B.,
 RA Antonarakis S.E., Bonne-Tamir B.;
 RT "Refined localization of autosomal recessive nonsyndromic deafness
 RL DFNB10 locus using 34 novel microsatellite markers, genomic
 structure, and exclusion of six known genes in the region.";
 RL Genomics 68:22-29 (2000).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RX MEDLINE=21192304; PubMed=11279031;
 RA Porach-Oezuermez M., Langmann T., Heimerl S., Borukova H.,
 RA Kaninski W.E., Drobnik W., Honer C., Schumacher C., Schmitz G.;
 RT "The zinc finger protein 202 (ZNF202) is a transcriptional repressor
 of ATP binding cassette transporter A1 (ABCA1) and ABCG1 gene
 expression and a modulator of cellular lipid efflux.";
 RL J. Biol. Chem. 276:12427-12433 (2001).
 RN [5]
 RP SEQUENCE FROM N.A. (ISOFORMS 2; 3; 4; 5; 6 AND 7).
 RX MEDLINE=21092576; PubMed=11162488;
 RA Lorkowski S., Rust S., Engel T., Jung E., Tegelkamp K., Galinski E.A.,
 RA Assmann G., Cullen P.;
 RT "Genomic sequence and structure of the human ABCG1 (ABC8) gene.";
 RL Biochem. Biophys. Res. Commun. 280:121-131 (2001).
 RN [6]
 RP SEQUENCE OF 33-678 FROM N.A.
 RC TISSUE=Fetal brain;
 RX MEDLINE=97186700; PubMed=9034316;
 RA Croop J.M., Tiller G.E., Fletcher J.A., Lux M.L., Raab E.,
 RA Goldenson D., Arciniegas S., Son D., Wu R.;
 RT "Isolation and characterization of a mammalian homologue of the
 RL Drosophila white gene.";
 RN Gene 185:77-85 (1997).
 RN [7]
 RP INDUCTION, AND PROBABLE FUNCTION.
 RX MEDLINE=20261604; PubMed=10799558;
 RA Venkateswaran A., Repa J.J., Lobaccaro J.-M.A., Bronson A.,
 RA Mangelsdorf D.J., Edwards P.A.;
 RT "Human white/murine ABC8 mRNA levels are highly induced in
 lipid-loaded macrophages. A transcriptional role for specific

oxysterols.";
 J. Biol. Chem. 275:14700-14707(2000).
 [8]
 INDUCTION, AND PROBABLE FUNCTION.
 MEDLINE=20105556; PubMed=10639163;
 Klucken J., Buechler C., Orso E., Kaminski W.E.,
 Porsch-Oezuermez M., Liebisch G., Kapinsky M., Diederich W.,
 Drobnik W., Dean M., Allikmets R., Schmitz G.;
 "ABCG1 (ABCG8), the human homolog of the Drosophila white gene, is a
 regulator of macrophage cholesterol and phospholipid transport.";
 Proc. Natl. Acad. Sci. U.S.A. 97:817-822(2000).
 [9]
 REVIEW.
 MEDLINE=21474438; PubMed=11590207;
 Schmitz G., Langmann T., Heimerl S.;
 "Role of ABCG1 and other ABCG family members in lipid metabolism.";
 J. Lipid Res. 42:1513-1520(2001).
 CC -!- FUNCTION: Transporter involved in macrophage lipid homeostasis. Is
 an active component of the macrophage lipid export complex. Could
 also be involved in intracellular lipid transport processes. The
 role in cellular lipid homeostasis may not be limited to
 macrophages.
 CC -!- SUBUNIT: May form heterodimers with several heterologous partners
 of the ABCG subfamily.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Predominantly
 localized in the intracellular compartments mainly associated with
 the endoplasmic reticulum (ER) and Golgi membranes.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=7;
 CC Comment=Additional isoforms seem to exist;
 CC Name=1;
 CC IsoId=P45844-1; Sequence=Displayed;
 CC Name=2; Synonyms=J;
 CC IsoId=P45844-2; Sequence=VSP_000047, VSP_000051;
 CC Name=3; Synonyms=ABDE;
 CC IsoId=P45844-3; Sequence=VSP_000048, VSP_000051;
 CC Name=4; Synonyms=G;
 CC IsoId=P45844-4; Sequence=VSP_000051;
 CC Name=5; Synonyms=F;
 CC IsoId=P45844-5; Sequence=VSP_000049, VSP_000051;
 CC Name=6; Synonyms=Hi;
 CC IsoId=P45844-6; Sequence=VSP_000046, VSP_000051;
 CC Name=7; Synonyms=C;
 CC IsoId=P45844-7; Sequence=VSP_000050, VSP_000051;
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN SEVERAL TISSUES.
 CC -!- INDUCTION: Strongly induced in monocyte-derived macrophages during
 cholesterol influx. Conversely, mRNA and protein expression are
 suppressed by lipid efflux. Induction is mediated by the liver X
 receptor/retinoid X receptor (LXR/RXR) pathway.
 CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCG (WHITE)
 SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X91249; CAAG2631.1; ALT INIT.
 CC EMBL; AP001746; BAA95530.1; ALT INIT.
 CC EMBL; AB038161; BAB13728.2; ALT INIT.
 CC EMBL; AJ289137; CAC00730.1; ALT INIT.
 CC EMBL; AJ289138; CAC00730.1; JOINED.
 CC EMBL; AJ289139; CAC00730.1; JOINED.
 CC EMBL; AJ289140; CAC00730.1; JOINED.
 CC EMBL; AJ289141; CAC00730.1; JOINED.
 CC EMBL; AJ289142; CAC00730.1; JOINED.
 CC EMBL; AJ289143; CAC00730.1; JOINED.
 CC EMBL; AJ289144; CAC00730.1; JOINED.
 CC EMBL; AJ289145; CAC00730.1; JOINED.
 CC EMBL; AJ289146; CAC00730.1; JOINED.

DR EMBL; AJ289147; CAC00730.1; JOINED.
 DR EMBL; AJ289148; CAC00730.1; JOINED.
 DR EMBL; AJ289149; CAC00730.1; JOINED.
 DR EMBL; AJ289150; CAC00730.1; JOINED.
 DR EMBL; AJ289151; CAC00730.1; JOINED.
 DR EMBL; AF323658; AAK28836.1;
 DR EMBL; AF323644; AAK28836.1; JOINED.
 DR EMBL; AF323645; AAK28836.1; JOINED.
 DR EMBL; AF323646; AAK28836.1; JOINED.
 DR EMBL; AF323647; AAK28836.1; JOINED.
 DR EMBL; AF323648; AAK28836.1; JOINED.
 DR EMBL; AF323649; AAK28836.1; JOINED.
 DR EMBL; AF323650; AAK28836.1; JOINED.
 DR EMBL; AF323651; AAK28836.1; JOINED.
 DR EMBL; AF323652; AAK28836.1; JOINED.
 DR EMBL; AF323653; AAK28836.1; JOINED.
 DR EMBL; AF323654; AAK28836.1; JOINED.
 DR EMBL; AF323655; AAK28836.1; JOINED.
 DR EMBL; AF323656; AAK28836.1; JOINED.
 DR EMBL; AF323657; AAK28836.1; JOINED.
 DR EMBL; AF323664; AAK28842.1;
 DR EMBL; AF323665; AAK28833.1;
 DR EMBL; AF323640; AAK28833.1; JOINED.
 DR EMBL; AF323645; AAK28833.1; JOINED.
 DR EMBL; AF323646; AAK28833.1; JOINED.
 DR EMBL; AF323647; AAK28833.1; JOINED.
 DR EMBL; AF323648; AAK28833.1; JOINED.
 DR EMBL; AF323649; AAK28833.1; JOINED.
 DR EMBL; AF323650; AAK28833.1; JOINED.
 DR EMBL; AF323651; AAK28833.1; JOINED.
 DR EMBL; AF323652; AAK28833.1; JOINED.
 DR EMBL; AF323653; AAK28833.1; JOINED.
 DR EMBL; AF323654; AAK28833.1; JOINED.
 DR EMBL; AF323655; AAK28833.1; JOINED.
 DR EMBL; AF323656; AAK28833.1; JOINED.
 DR EMBL; AF323657; AAK28833.1; JOINED.
 DR EMBL; AF323660; AAK28838.1;
 DR EMBL; AF323663; AAK28841.1; ALT_INIT.
 DR EMBL; AF323658; AAK28835.1; JOINED.
 DR EMBL; AF323642; AAK28835.1; JOINED.
 DR EMBL; AF323645; AAK28835.1; JOINED.
 DR EMBL; AF323646; AAK28835.1; JOINED.
 DR EMBL; AF323647; AAK28835.1; JOINED.
 DR EMBL; AF323648; AAK28835.1; JOINED.
 DR EMBL; AF323649; AAK28835.1; JOINED.
 Query Match 14.7%; Score 113; DB 1; Length 678;
 Best Local Similarity 24.1%; Pred. No. 0.0076;
 Matches 32; Conservative 28; Mismatches 65; Indels 8; Gaps 2;
 Qy 10 ASSMALATATGQSVVSVATLMTATFAVFMMLFSGLLVNLRTIGPWLWLOVFSIPRYGFT 69
 Db 548 AQSGLLGAASTSLQVATFVGPVTAIPVLLFSGFVSFDITYLYQMWSYISVRYGFE 607
 Qy 70 ALQVNEFLGFCFGFNVDNSTCVNSYAICTGNEYLINQGIELSPWGLWKNHVALACMI 129
 Db 608 GVILSIY-----GLREDLHCDIDETCHQKSEAILRE-LDVENAKLYLDLFIYVIGFF 659
 Qy 130 IIFLTIAVYKLLF 142
 Db 660 ISLRLLIAYFVLR 672
 RESULT 7
 YN99 YEAST
 ID YN99 YEAST STANDARD; PRT; 1333 AA.
 AC P53756;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Probable ATP-dependent transporter YNR070W.
 GN YNR070W OR N3568.
 OS Saccharomyces cerevisiae (Baker's yeast).

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; U34920; ABA47738.1; -
DR ENBL; Z48745; CAA88636.1; -
DR MBL; AF323659; AAK27442.1; -
DR MGD; MGI:107704; Abcg1.
DR InterPro; IPR003593; AAA ATPase.
DR InterPro; IPR003439; ABC transporter.
DR InterPro; IPR005284; Pigment_permease.
DR Pfam; PF00005; ABC tran; 1.
DR ProDom; PD000006; ABC transporter; 1.
DR SMART; SM00382; AAA; 1.
DR TIGRfam; TIGR00955; 3a01204; 1.
DR PROSITE; PS00211; ABC TRANSPORTER 1; 1.
DR PROSITE; PS00893; ABC TRANSPORTER 2; 1.
KW ATP-binding; Transmembrane; Transport.
FT DOMAIN 1 414 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 415 433 POTENTIAL.
FT DOMAIN 434 444 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 445 465 POTENTIAL.
FT DOMAIN 466 494 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 495 513 POTENTIAL.
FT DOMAIN 514 521 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 522 543 POTENTIAL.
FT DOMAIN 544 555 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 556 574 POTENTIAL.
FT DOMAIN 575 637 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 638 657 POTENTIAL.
FT DOMAIN 658 666 CYTOPLASMIC (POTENTIAL).
FT NP_BIND 118 125 ATP (POTENTIAL).
SQ SEQUENCE 666 AA; 74033 MW; EDDC6AFED4395086 CRC64;

Query Match 14.3%; Score 110; DB 1; Length 666;
Best Local Similarity 37.3%; Pred. No. 0.014;
Matches 22; Conservative 11; Mismatches 26; Indels 0; Gaps 0;

QY 10 ASMALAIATGQSVSVAILMTIAFVFMFLSGLLVNRITGPMLSWLYQPSIPRYG 68
DB 536 AOSGLLLGAASTLQVATFVGVTAPVLLFSGFVSDTIPAVLQWMSYISVRYGF 594

RESULT 9
CDRI_CANAL
ID CDRI_CANAL STANDARD; PRT; 1501 AA.
AC P43071;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Multidrug resistance protein CDRI.
GN CDRI.

OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5476;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 64385 / 1001;
RX MEDLINE=95339406; PubMed=7614555;
RA Prasad R., de Wergifosse P., Balzi E., Goffeau A.;
RT "Molecular cloning and characterization of a novel gene of Candida
RT albicans, CDRI, conferring multiple resistance to drugs and
RT antifungals."
RT Curr. Genet. 27:320-329 (1995).
CC -!- FUNCTION: TRANSPORTER, WHOSE PHYSIOLOGICAL FUNCTION IS NOT YET
CC ESTABLISHED. CONFERS RESISTANCE TO THE CHEMICAL CYCLOHEXIMIDE.
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. PDR5 SUBFAMILY.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; X77589; CAA54692.1; -
DR PIR; S57198; S57198.
DR InterPro; IPR003593; AAA ATPase.
DR InterPro; IPR003439; ABC transporter.
DR InterPro; IPR005285; PDR.
DR Pfam; PF00005; ABC tran; 2.
DR ProDom; PD000006; ABC transporter; 2.
DR SMART; SM00382; AAA; 2.
DR TIGRfam; TIGR00956; 3a01205; 1.
DR PROSITE; PS00211; ABC TRANSPORTER 1; 1.
DR PROSITE; PS00893; ABC TRANSPORTER 2; 2.
KW ATP-binding; Transmembrane; Glycoprotein; Transport; Repeat.
FT DOMAIN 1 513 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 514 534 POTENTIAL.
FT TRANSMEM 549 569 POTENTIAL.
FT TRANSMEM 598 618 POTENTIAL.
FT TRANSMEM 623 643 POTENTIAL.
FT TRANSMEM 655 675 POTENTIAL.
FT TRANSMEM 765 785 POTENTIAL.
FT DOMAIN 786 1195 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 1196 1216 POTENTIAL.
FT TRANSMEM 1230 1250 POTENTIAL.
FT TRANSMEM 1281 1301 POTENTIAL.
FT TRANSMEM 1315 1335 POTENTIAL.
FT TRANSMEM 1357 1377 POTENTIAL.
FT TRANSMEM 1467 1487 POTENTIAL.
FT NP_BIND 895 902 ATP (POTENTIAL).
SQ SEQUENCE 1501 AA; 169937 MW; BB144A0BAD7ED233 CRC64;

Query Match 14.2%; Score 109.5; DB 1; Length 1501;
Best Local Similarity 20.7%; Pred. No. 0.032;
Matches 34; Conservative 33; Mismatches 70; Indels 27; Gaps 5;

QY 8 YTSALMAIATGQSVSVAILMTIAFVFMFLSGLLVNRITGPMLSWLYQPSIPRYG 67
DB 637 FVMHLPFRISGAVSTISGAMTPATVLLAMVITGVFPTPSMLGMSRWINYINPVGV 696
QY 68 FTALQYNFLEQGF-C-----PGF-NVTDSNCTVNSYAICTGNEYILNQIELSPWGL 118
DB 697 FESLMVNEFHGRFQCAQYVPSGFGYENISKNQVCTAGVSPGNEVMWVGNYLAGAYQY 756
QY 119 WKNH-----VALACMIIFLTIAYLKL-----IFLK 144
DB 757 YNSHKWRNLGITIGFAVFFLAIIYIALTEFNKGAMQKGEIVFLK 800

RESULT 10
SNQ2_YEAST
ID SNQ2_YEAST STANDARD; PRT; 1501 AA.
AC P32568;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE SNQ2 protein.
GN SNQ2 OR YDR011W OR YDR119.16.
OS Saccharomycetes cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93173094; PubMed=8437567;
RA Servos J., Hease E., Brendel M.;
RT "Gene SNQ2 of Saccharomycetes cerevisiae, which confers resistance to
RT 4-nitroquinoline-N-oxide and other chemicals, encodes a 169 kDa
RT protein homologous to ATP-dependent permeases."
RT Mol. Gen. Genet. 236:214-218 (1993).

10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

Query Match 14.2%; Score 109.5; DB 1; Length 1501;
Best Local Similarity 23.6%; Pred. No. 0.032;
Matches 35; Conservative 30; Mismatches 56; Indels 27; Gaps 7;

Query Match 14.1%; Score 108.5; DB 1; Length 598;
Best Local Similarity 27.8%; Pred. No. 0.017;
Matches 40; Conservative 24; Mismatches 65; Indels 15; Gaps 4;

Query Match 14.2%; Score 109.5; DB 1; Length 1501;
Best Local Similarity 23.6%; Pred. No. 0.032;
Matches 35; Conservative 30; Mismatches 56; Indels 27; Gaps 7;

Query Match 14.1%; Score 108.5; DB 1; Length 598;
Best Local Similarity 27.8%; Pred. No. 0.017;
Matches 40; Conservative 24; Mismatches 65; Indels 15; Gaps 4;

RESULT 14

PDRA	YEAST	STANDARD;	PRT; 1564 AA.
ID	PDRA_YEAST		
AC	PS1533;		
DT	01-OCT-1996 (Rel. 34, Created)		
DT	01-OCT-1996 (Rel. 34, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	ATP-dependent permease PDR10.		
GN	PDR10 OR YOR328W.		
OS	Saccharomyces cerevisiae (Baker's yeast).		
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;		
OC	Saccharomycetales; Saccharomycetaceae; Saccharomyces.		
NCBI_Taxid=4932;			
OX	[1]		
RN	SEQUENCE FROM N.A.		
RP			
RX	MEDLINE=97051586; PubMed=8996263;		
RA	Parle-McDermott A.G., Hand N.J., Goulding S.G., Wolfe K.H.;		
RT	"Sequence of 29 kb around the PDR10 locus on the right arm of		
RT	Saccharomyces cerevisiae chromosome XV: similarity to part of		
RT	chromosome I.";		
RL	Yeast 12.999-1004(1996).		
CC	-1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).		
CC	-1- SIMILARITY: Belongs to the ABC transporter family.		
CC			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announcement		
CC	or send an email to license@isb-sib.ch).		
CC			
EMBL; Z49821; CAA89975.1; -			
EMBL; Z75236; CAA99649.1; -			
EMBL; Z75237; CAA99651.1; -			
PIR; S55517; S55517.			
DR	GSD; S0005855; PDR10.		
DR	GQ; GQ:0006855; P:multidrug transport; IEP.		
DR	InterPro; IPR003593; AAA_ATPase.		
DR	InterPro; IPR003439; ABC_transporter.		
DR	InterPro; IPR005285; PDR-		
DR	Pfam; PF00005; ABC_tran; 2.		
DR	ProDom; PD000006; ABC_transporter; 2.		
DR	SMART; SM00382; AAA; 2.		
DR	TIGRFAMs; TIGR00956; 3a01205; 1.		
DR	PROSITE; PS00211; ABC_TRANSPORTER_1; 1.		
DR	PROSITE; PS50893; ABC_TRANSPORTER_2; 2.		
KW	ATP-binding; Transmembrane; Glycoprotein; Transport.		
FT	DOMAIN 1 587		
FT	CYTOPLASMIC (POTENTIAL).		
FT	POTENTIAL.		
FT	TRANSMEM 588 608		
FT			

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; Z38113; CAA86236.1; -;
DR EMBL; Z46881; CAA86980.1; -;
DR PIR; S48442; S48442.
DR SGD; S0001275; PDR11.
DR GO; GO:0015918; P:sterol transport; IMP.
DR InterPro; IPR003593; AAA ATPase.
DR InterPro; IPR003439; ABC transporter.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transporter; 2.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 2.
KW ATP-binding; Transmembrane; Glycoprotein; Transport.
FT INIT MET 0 0
FT DOMAIN 1 387 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 388 408 POTENTIAL.
FT TRANSMEM 418 438 POTENTIAL.
FT TRANSMEM 471 491 POTENTIAL.
FT TRANSMEM 494 514 POTENTIAL.
FT TRANSMEM 524 544 POTENTIAL.
FT TRANSMEM 536 556 POTENTIAL.
FT DOMAIN 657 1089 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 1090 1110 POTENTIAL.
FT TRANSMEM 1117 1137 POTENTIAL.
FT TRANSMEM 1175 1195 POTENTIAL.
FT TRANSMEM 1204 1224 POTENTIAL.
FT TRANSMEM 1230 1250 POTENTIAL.
FT TRANSMEM 1355 1375 POTENTIAL.
FT DOMAIN 1376 1410 CYTOPLASMIC (POTENTIAL).
FT NP BIND 781 788 ATP (POTENTIAL).
FT DOMAIN 707 712 POLY-SER.
FT DOMAIN 1045 1048 POLY-LEU.
FT CARBOHYD 594 594 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1288 1288 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1323 1323 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1345 1345 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1410 AA; 160405 MW; 93C9399A5CD114C3 CRC64;

Query Match 12.9%; Score 99.5; DB 1; Length 1410;
Best Local Similarity 19.9%; Pred. No. 0.24; Mismatches 59; Indels 35; Gaps 7;
Matches 31; Conservative 31; Mismatches 59; Indels 35; Gaps 7;
QY 24 VSVATLLMTAFVFMFLPSGLLVNLRITGFWLSWLYQFSIPRYGFTALQYNEFLG----- 78
DB 522 LSMANLLAGILLALAMAYASYVYMKDHPFWFIWAIYLNPAFAMEATILSNELFNKLDLC 581
QY 79 -QEFPCGFNVTONSTCVNSYAICT-----GNEYL-----INOGIELSPWGLWKNH---V 123
DB 582 HESIIPRGEYDNIISF--SHKACAWGATLGNDRDYKSLKLYTHVWRNFGIII 639
QY 124 ALACMIII--FLTIAYLKLIIF-----LKKY 146
DB 640 GFLCFFLCSLAAEYITELFTRENLRLWNLYKRY 675

Search completed: August 1, 2003, 18:48:37
Job time : 7.49875 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 1, 2003, 18:38:41 ; Search time 24.5012 Seconds
(without alignments)
1257.181 Million cell updates/sec

Title: US-09-856-927-2

Perfect score: 3350

Sequence: 1 MSSSNVEFIPVSGQNTNGF.....MIVIFLTAYLKLFLKKYS 655

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3344	99.8	655	1 ABG2 HUMAN	Q9unq0 homo sapien
2	1786.5	53.3	650	1 ABG3 MOUSE	Q99p81 mus musculus
3	835.5	24.9	1049	1 ADP1 YEAST	P25371 saccharomyc
4	803	24.0	687	1 WHIT DROME	P10090 drosophila
5	762.5	22.8	679	1 WHIT CERCA	Q17320 ceratilis c
6	748.5	22.3	677	1 WHIT LUCCU	Q05360 lucilia cup
7	745	22.2	695	1 WHIT ANOGA	Q27256 anopheles g
8	730.5	21.8	709	1 WHIT ANOAL	Q16928 anopheles a
9	712.5	21.3	666	1 ABG1 MOUSE	Q64343 mus musculus
10	704.5	21.0	678	1 ABG1 HUMAN	P45944 homo sapien
11	702	21.0	646	1 ABG4 HUMAN	Q9h172 homo sapien
12	690.5	20.6	652	1 ABG5 MOUSE	Q99pe8 mus musculus
13	678.5	20.3	652	1 ABG5 RAT	Q99pe7 rattus norv
14	676.5	20.2	598	1 YPC3 CAEL	Q11180 caenorhabdi
15	674.5	20.1	651	1 ABG5 HUMAN	Q9h222 homo sapien
16	671	20.0	672	1 ABG8 RAT	P58428 rattus norv
17	664	19.8	673	1 ABG8 MOUSE	Q9dbm0 mus musculus
18	653.5	19.5	1294	1 YOH5 YEAST	Q08234 saccharomyc
19	649.5	19.4	666	1 SRT DROME	P45843 drosophila
20	637.5	19.0	673	1 ABG8 HUMAN	Q9h221 homo sapien
21	618.5	18.5	610	1 YQSC CAEL	Q09466 caenorhabdi
22	605.5	18.1	1501	1 SNQ2 YEAST	P32568 saccharomyc
23	603	18.0	1530	1 BFR1 SCHPO	P41820 schizosacch
24	601.5	18.0	1499	1 CDR2 CANAL	P78595 candida alb
25	596	17.8	1511	1 PDR5 YEAST	P33302 saccharomyc
26	588	17.6	1564	1 PDR4 YEAST	P51533 saccharomyc
27	579	17.3	1333	1 YN99 YEAST	P53756 saccharomyc
28	569	17.0	1529	1 PDRF YEAST	Q04182 saccharomyc
29	537	16.0	1501	1 CDR3 CANAL	O42690 candida alb
30	530.5	15.8	1501	1 CDR1 CANAL	P43071 candida alb
31	527.5	15.7	1511	1 PDRC YEAST	Q02785 saccharomyc
32	501	15.0	1490	1 CDR4 CANAL	O74676 candida alb
33	483.5	14.4	1410	1 PDRB YEAST	P40550 saccharomyc

RESULT 1
ABG2 HUMAN
ID ABG2 HUMAN STANDARD; PRT; 655 AA.
AC Q9UNQ0; Q95344; Q9BY73; Q9NUS0;
DT 16-OCT-2001 Rel. 40, Created)
DT 28-FEB-2003 Rel. 41, Last sequence update)
DT 28-FEB-2003 Rel. 41, Last annotation update)
DE ATP-binding cassette, sub-family G, member 2 (Placenta-specific ATP-binding cassette transporter) (Breast cancer resistance protein).
GN ABCG2 OR ABCP OR BCRP OR BCRP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RF MEDLINE=99065313; PubMed=9850061;
RA Allikmets R., Schriml L.M., Hutchinson A., Romano-Spica V., Dean M.;
RT "A human placenta-specific ATP-binding cassette gene (ABCP) on chromosome 4q22 that is involved in multidrug resistance.";
RL Cancer Res. 58:5337-5339(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast cancer;
RF MEDLINE=99080071; PubMed=9861027;
RA Doyle L.A., Yang W., Abruzzo L.V., Krogmann T., Gao Y., Rishi A.K., Ross D.D.;
RT "A multidrug resistance transporter from human MCF-7 breast cancer cells";
RL Proc. Natl. Acad. Sci. U.S.A. 95:15665-15670(1998).
RN [3]
RP ERBATUM.
RA Doyle L.A., Yang W., Abruzzo L.V., Krogmann T., Gao Y., Rishi A.K., Ross D.D.;
RT "Breast cancer resistance protein constitutes a 140-kDa complex as a homodimer";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 198-655 FROM N.A.
RC TISSUE=Placenta;
RF MEDLINE=99080071; PubMed=9861027;
RA Isogai T., Oca T., Hayaashi K., Sugiyama T., Otauki T., Suzuki Y., Nishikawa T., Negai K., Sugano S., Shiratori A., Sudo H., Sugawara M., Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuho Y., Ninomiya K., Iwayanagi T.;
RT "NEDO human cDNA sequencing project";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [6]

ALIGNMENTS

34 464 13.9 675 1 BROW DROME P12428 drosophila
35 445 13.3 668 1 BROW DROVI Q24739 drosophila
36 237 7.1 355 1 CYSA SYN3 P74548 synchocyst
37 236.5 7.1 1704 1 ABC3 HUMAN Q99758 homo sapien
38 236 7.0 246 1 NATA_BACSU P46903 bacillus au
39 236 7.0 371 1 MALK_ECOLI P02914 escherichia
40 236 7.0 1321 1 AB11 HUMAN O95342 homo sapien
41 235.5 7.0 236 1 LIVF_ARCFU O28882 archaeglob
42 232.5 6.9 607 1 HEPA_ANASP P22638 anabaena sp
43 227 6.8 1302 1 DROME Q00449 drosophila
44 225.5 6.7 576 1 CYDC_HAETN P45081 haemophilus
45 224 6.7 347 1 NODI_RHIGA P50332 rhizobium g

RP REVIEW.
RX MEDLINE=2147438; PubMed=11590207;
RA Schmitz G., Langmann T., Heimerl S.;
RT "Role of ABCG1 and other ABCG family members in lipid metabolism."; J. Lipid Res. 42:1513-1520(2001).
RL J. Lipid Res. 42:1513-1520(2001).
CC -!- FUNCTION: XENOBIOTIC TRANSPORTER THAT APPEARS TO PLAY A MAJOR ROLE IN THE MULTIDRUG RESISTANCE PHENOTYPE OF A SPECIFIC MCF-7 BREAST CANCER CELL LINE. WHEN OVEREXPRESSED, THE TRANSFECTED CELLS BECOME RESISTANT TO MITOXANTHONE, DAUNORUBICIN AND DOXORUBICIN, DISPLAY DIMINISHED INTRACELLULAR ACCUMULATION OF DAUNORUBICIN, AND MANIFEST AN ATP-DEPENDENT INCREASE IN THE EFFLUX OF RHODAMINE 123.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCG (WHITE) SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF103796; AAC09188.1; -;
DR EMBL; AF098951; AAC97367.1; -;
DR EMBL; AB056867; BAB39212.1; -;
DR EMBL; AK002040; BAA92050.1; -;
DR Genew; HGNC:74; ABCG2.
DR MIM; 603756; -;
DR GO; GO:0016021; C:integral to membrane; TAS.
DR GO; GO:0005524; F:ATP binding activity; TAS.
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; TAS.
DR GO; GO:0005215; F:transporter activity; TAS.
DR GO; GO:0008559; F:xenobiotic-transporting ATPase activity; TAS.
DR GO; GO:0009315; F:drug resistance; TAS.
DR GO; GO:0006832; P:small molecule transport; TAS.
DR InterPro; IPR003593; AAA ATPase.
DR InterPro; IPR003439; ABC transporter.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; FALSE_NEG.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
KW ATP-binding; Transmembrane; Transport.
FT DOMAIN 1 395 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 396 416 POTENTIAL.
FT DOMAIN 417 428 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 429 449 POTENTIAL.
FT DOMAIN 450 477 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 478 498 POTENTIAL.
FT DOMAIN 499 506 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 507 527 POTENTIAL.
FT DOMAIN 528 535 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 536 556 POTENTIAL.
FT DOMAIN 557 630 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 631 651 POTENTIAL.
FT DOMAIN 652 655 CYTOPLASMIC (POTENTIAL).
FT NP_BIND 80 87 ATP (POTENTIAL).
FT CARBOHYD 418 418 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 557 557 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 596 596 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 24 24 V -> A (IN REF. 2 AND 4).
FT CONFLICT 166 166 E -> Q (IN REF. 2 AND 4).
FT CONFLICT 208 208 F -> S (IN REF. 1).
FT CONFLICT 208 208 F -> S (IN REF. 1).
FT CONFLICT 315 316 MISSING (IN REF. 5).
FT CONFLICT 482 482 R -> T (IN REF. 2).
SQ SEQUENCE 655 AA; 72343 MW; 89A6D3511DC5CEE0 CRC64;
Query Match 99.8%; Score 3344; DB 1; Length 655;
Best Local Similarity 99.8%; Pred.No. 3.3e-221;
Matches 654; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
1 MSSSNVEVFPVQSGNTNGFPATVNSDLKAFTEGAVLSFHNICRYRVLKSGFLPCRKPYE 60

Db 1 MSSSNVEVFPVQSGNTNGFPATVNSDLKAFTEGAVLSFHNICRYRVLKSGFLPCRKPYE 60
Qy 61 KEILSNINGIMKPGNALIGPTGGKSSLLDVLAAARKOPSGLSGVDLINGAPRANPKCN 120
Db 61 KEILSNINGIMKPGNALIGPTGGKSSLLDVLAAARKOPSGLSGVDLINGAPRANPKCN 120
Qy 121 SGYVQDDVVGTLTVRENLOFSAALRLATTWNHEKNERINRVTEELGLDKVADSKVGT 180
Db 121 SGYVQDDVVGTLTVRENLOFSAALRLATTWNHEKNERINRVTEELGLDKVADSKVGT 180
Qy 181 QFIRVSGGERKRTSIGMELITDPSILSDLEPTTGLDSTANAVALLKRMKSKQGRITIF 240
Db 181 QFIRVSGGERKRTSIGMELITDPSILSDLEPTTGLDSTANAVALLKRMKSKQGRITIF 240
Qy 241 SIHQPRYSIFKLFDSLTLTLLASGRLMFHGPAQEALGYFESAGYHCEAYNNPADFFLDIING 300
Db 241 SIHQPRYSIFKLFDSLTLTLLASGRLMFHGPAQEALGYFESAGYHCEAYNNPADFFLDIING 300
Qy 301 DSTAVALNREDDFKATEIIEBPSKODKPLIEKLAETIYVNSFVKETKAELHQLSGGKKKK 360
Db 301 DSTAVALNREDDFKATEIIEBPSKODKPLIEKLAETIYVNSFVKETKAELHQLSGGKKKK 360
Qy 361 ITVFKEISYTTSFCHQLRWKRSFKNLLGNPQASIAQIIVTVVLGLVIGAIYFGLKNDK 420
Db 361 ITVFKEISYTTSFCHQLRWKRSFKNLLGNPQASIAQIIVTVVLGLVIGAIYFGLKNDK 420
Qy 421 TGIQNRAGVLFELTNQCFSSVAVELFVVEKKLFIHEIYISGYRVSSYFGLKLSDDL 480
Db 421 TGIQNRAGVLFELTNQCFSSVAVELFVVEKKLFIHEIYISGYRVSSYFGLKLSDDL 480
Qy 481 MRMLPSIIFTCIVYFMLGLKPKADAFVVMFTLMWVAYSASSMALAIAGQSVSVATLL 540
Db 481 MRMLPSIIFTCIVYFMLGLKPKADAFVVMFTLMWVAYSASSMALAIAGQSVSVATLL 540
Qy 541 MTICFVFMWIFSGLLVNLTTIASWLSWLYQSFIPRYGFTALQHNFLQNGFCPLNATGN 600
Db 541 MTICFVFMWIFSGLLVNLTTIASWLSWLYQSFIPRYGFTALQHNFLQNGFCPLNATGN 600
Qy 601 NPCNATCTGEEYLVKQIDILSPWGLWKNHVALACMIVIFLTIAVLLKLLFKKYS 655
Db 601 NPCNATCTGEEYLVKQIDILSPWGLWKNHVALACMIVIFLTIAVLLKLLFKKYS 655
RESULT 2
ABG3_MOUSE
ID ABG3_MOUSE STANDARD; PRY 650 AA.
AC Q99P81;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE ATP-binding cassette, sub-family G, member 3.
GN ABCG3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Spleen;
RX MEDLINE=21030753; PubMed=11178751;
RA Mickley L., Jain P., Miyake K., Schriml L.M., Rao K., Fojo T.,
RT "An ATP-binding cassette gene (ABCG3) closely related to the multidrug
RT transporter ABCG2 (MXR/ABCP) has an unusual ATP-binding domain.";
RL Mamm. Genome 12:86-88(2001).
CC -!- SUBUNIT: May dimerize with another subunit to form a functional
CC transporter.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -!- TISSUE SPECIFICITY: Highest levels of expression in thymus and
CC spleen. Detected in lung and small intestine.
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCG (WHITE)
CC SUBFAMILY.

RA Benos P.V., Gatt M.K., Ashburner M., Murphy L., Harris D.,
RA Barrell B.G., Ferraz C., Vidal S., Brun C., Demailles J., Cadieu E.,
RA Dreano S., Gloux S., Lelaure V., Mottier S., Galibert F., Borkova D.,
RA Minana B., Kafatos F.C., Louis C., Siden-Kiamos I., Bolehakov S.,
RA Papagiannakis G., Spanos L., Cox S., Madueno E., de Pablo F.,
RA Modolell J., Peter A., Schoettler P., Werner M., Mourikioti F.,
RA Beirato N., Dowe G., Schaefer U., Jaecle H., Bucheton A.,
RA Callister D.M., Campbell L.A., Darlamitsou A., Henderson N.S.,
RA McMillan P.J., Salles C., Tait E.A., Valenti P., Saunders R.D.C.,
RA Glover D.M.;
RT "From sequence to chromosome: the tip of the X chromosome of D.
RT melanogaster";
RT Science 287:2220-2222(2000).
RN [6]
RP SEQUENCE OF 224-331 FROM N.A.
RX MEDLINE=89339145; PubMed=2503416;
RA Tearle R.G., Belote J.M., McKeown M., Baker B.S., Howells A.J.;
RT "Cloning and characterization of the scarlet gene of Drosophila
RT melanogaster";
RL Genetics 122:595-606(1989).
CC -!- FUNCTION: PART OF A MEMBRANE-SPANNING PERMEASE SYSTEM NECESSARY
CC FOR THE TRANSPORT OF PIGMENT PRECURSORS INTO PIGMENT CELLS
CC RESPONSIBLE FOR EYE COLOR. WHITE DIMERIZE WITH BROWN FOR THE
CC TRANSPORT OF GUANINE AND WITH SCARLET FOR THE TRANSPORT OF
CC TRIPTOPHAN.
CC -!- SUBUNIT: HETERODIMER OF WHITE WITH EITHER BROWN OR SCARLET.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; X51749; CAA36038.1; -;
CC EMBL; X02974; CAA26716.1; -;
CC EMBL; AB028139; BAA78210.1; -;
CC EMBL; AE003425; AAF45826.1; -;
CC EMBL; AL133506; CAB65847.1; -;
CC EMBL; X76202; CAA53795.1; -;
CC PIR; S08635; FYFFW.
CC FlyBase; FBgn003996; w.
CC GO; GO:004888; F:transmembrane receptor activity; NAS.
CC GO; GO:0006727; P:omochrome biosynthesis; IMP.
CC InterPro; IPR003593; AAA ATPase.
CC InterPro; IPR003439; ABC transporter.
CC InterPro; IPR005284; Pigment_permease.
CC Pfam; PF00005; ABC_tran; 1.
CC ProDom; PD000006; ABC transporter; 1.
CC SMART; SM00382; AAA; 1.
CC TIGRfam; TIGR00955; 3a01204; 1.
CC PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
CC PROSITE; PS0893; ABC_TRANSPORTER_2; 1.
CC KW Pigment; ATP-binding; Transmembrane; Transport.
CC NP_BIND 130 137 ATP (BY SIMILARITY).
CC FT TRANSMEM 435 453 POTENTIAL.
CC FT TRANSMEM 455 485 POTENTIAL.
CC FT TRANSMEM 515 533 POTENTIAL.
CC FT TRANSMEM 542 563 POTENTIAL.
CC FT TRANSMEM 576 594 POTENTIAL.
CC FT TRANSMEM 659 678 POTENTIAL.
CC FT CONFLICT 25 29 GDGSA -> LIFEIPHCRTVAD (IN REF. 2 AND
CC FT 3).
CC FT L -> R (IN REF. 4 AND 5).
CC FT VQAQCPYNNPADFYVQVLAVVPGREIESRDIKIC ->
CC FT ITHLNSYPAMVPSVLPVTRTFTYRCWPLCPDGRSSPVI
CC FT GSPRYG (IN REF. 3).
CC SEQUENCE 687 AA; 75672 MW; 24FAD799DE0D396 CRC64;
SQ

Query Match

24.0%; Score 803; DB 1; Length 687;

Best Local Similarity 33.2%; Pred. No. 2.4e-47;
Matches 199; Conservative 127; Mismatches 233; Indels 40; Gaps 14;
QY 61 KEILSNINGIMKPG-LNAILLGPTGGKSSLLDLVAARKDPGSL-----SGDVLINGAPPA 115
DB 110 KHLKNVCGVAYPGELLAVMGSSGAGKTTLLNALAFR-SPOGIQVSPSGMRLNGQPVDA 168
QY 116 -NFKNSGYVVDVVMGTLTVRENLOPSAALRLATMTNHEKNERINRVIEELGLDKVA 174
DB 169 KEMQRCAYVQDDLFISGLTAREHLIFQAMVRPHLTQRVARVDVQIELSLKQ 228
QY 175 DSKVGTQ-FIRGVSGGKRTSIGHELITDPSILSLDEPTTGLDSSSTANAVLLLRMSK 233
DB 229 HTIIGVPRVKGLSGGERKRLAFASEALTDPLLCDEPTSGLDSTFSAHVQVVLKLSQ 288
QY 234 QGRITPSIHOPRYSIPKLPDLSTLLASGRLMWHPAQEALGYFESAGYCEAYNNPADF 293
DB 289 KGTIVILTIHQPSSSELPFDKILLMAEGRVAFLTGTPSEAVDFSYVGAQCPTNNPADF 348
QY 294 FLDIINGDSTAVALNRBEDFKATBIIEPSKQDKPLIEKLAIEIYVNSSPYKETKAEHLHLS 353
DB 349 YVQVL-----AVVPGREIESR-----DRIAKICDNFAISKVAR-DMEQLL 387
QY 354 GGEKKKITVPEKISYT--TSFCHQLRWISKRSFKNLGNPQASIAQIIVTVLGLVIGA 411
DB 388 ATKLEKPLEQEPENGYTYKATFMQFRAVLWRSVLSVLEPLLVKRVLIQTMTVAIIQL 447
QY 412 IYFGLKNDSTGIONRAGVLPVLTNNQCFSSVSA-VELFVVEKKLPFIEHIVSYGVRSVYP 470
DB 448 IFLQOQLTQVGMVINGAIFLFTNMTFQNVFATINVTSELPMRARSRLRYCDTYP 507
QY 471 LGKLLSLLPMRMPLPSIIFTIVYFMGLGKPKADAFVMMFTLMVAVSASSMALAIAG 530
DB 508 LGKTIAB-LPLFLTVPLVFTALAYPMIGLRAGLVHFFNCLALVTLVANVSTSPGLVLSA 566
QY 531 QSVSVATLMTICFVFMIMPISGLLVNLTITIASWLSWLYQYSPISRYGTALQHNEFLQN 590
DB 567 SSSTSMALSGPPVPIIFLFGGFLNSGSPVYLVKWSLYLSWFRYANEGELLINQWADVE 626
QY 591 FCPG-LNATGNPCNYATCTGEEVLVKQIDLSWGLWKNHVALACMIVILTTAYLKL 648
DB 627 --PGEISCTSN-----TTCPSGKVILETNLFNSAADLPDYVGLAILIVSPRVLAYLAL 679

RESULT 5
WHIT_CERCA
ID WHIT_CERCA STANDARD; PRT; 679 AA.
AC Q17320;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE White protein.
GN W.
OS Ceratitis capitata (Mediterranean fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Tephritidae; Tephritidae; Ceratitis.
OX NCBI_TaxID=7213;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96123276; PubMed=8533095;
RA Zwiebel L.J., Saccone G., Zacharopoulos A., Besanaky N.J.,
RA Favia G., Collins F.H., Louis C., Kafatos F.C.;
RT "The white gene of Ceratitis capitata: a phenotypic marker for
RT germline transformation";
RL Science 270:2005-2007(1995).
CC -!- FUNCTION: MAY BE PART OF A MEMBRANE-SPANNING PERMEASE SYSTEM
CC NECESSARY FOR THE TRANSPORT OF PIGMENT PRECURSORS INTO PIGMENT
CC CELLS RESPONSIBLE FOR EYE COLOR.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch)

DR EMBL; X89933; CAA61998.1; -
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR005284; Pigment_permease.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR TIGRfams; TIGR00955; 3a01204; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
KW Pigment; ATP-binding; Transmembrane; Transport.
FT NP_BIND 121 128 ATP (BY SIMILARITY).
FT TRANSMEM 427 445 POTENTIAL.
FT TRANSMEM 457 477 POTENTIAL.
FT TRANSMEM 507 525 POTENTIAL.
FT TRANSMEM 534 555 POTENTIAL.
FT TRANSMEM 568 586 POTENTIAL.
FT TRANSMEM 651 670 POTENTIAL.
FT CARBOHYD 628 628 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 643 643 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 679 AA; 75145 MW; 3F9CB78A835C4CC CRC64;

Query Match 22.8%; Score 762.5; DB 1; Length 679;
Best Local Similarity 31.1%; Pred. No. 1.4e-44;
Matches 188; Conservative 130; Mismatches 237; Indels 49; Gaps 13;

QY 61 KEILSNINGIMKPG-LNALIGTGGKSSLLDVLARK-----DPSGLSGVDLNGAPR 113
DB 101 KELLKNDGAVAYPGELLAVMGSSGAKTLLNASAFRSGKVGQISPTIR---MLNGHPV 157
QY 114 PA-NFKCNISGVYQDDVNVGTLTRENLOFSAALRLATMTNHEKNERINRVIEELGDK 172
DB 158 DAKEMQARCAVYQDDLFGLSITAREHLIFQAVRMPRMTQKQVQVQVQVQVQVQV 217
QY 173 VADSKVGTG-FIRGVSGGKRTSIGMELITDPSILSDPTGLDSSANAVALLKRM 231
DB 218 CQNTLIGVGRVKGSLGGERKLAFASEALTDPLLCIDPTSGLDSEMAHVSQVQVQV 277
QY 232 SQQGTIIPSIHOPYSIFKLDLSLTSLASGLMFGPAQALGFESAGHYCEAYNPA 291
DB 278 SQGKKTIVLTTHQPSSELPFELDKILLMAEGRVAFGLTPEGAVDFFSYIGATCPTNYTPA 337
QY 292 DFFLDLINDGSTAVALNREED-----FKATEIIEPSKQDKPLIEKLAITYVNSFVKETK 346
DB 338 DFYVQVL-----AVVPGREVEDRVAKICDNFVGVKVSREMEQFQKLKVSNGFGKDE 392
QY 347 AELHQLSGGKKKKITVFKIEISYTTSFCHQLRWKSKRPNLLGNPOASIAQIIVTVVLG 406
DB 393 -----NEYTKASWFMQFRAVLWRSWLSVLKEPLLKVKVLLQTTWA 434
QY 407 LVIGAIYFGLKNDSTGIQNRAGVLPFLTNQCF-SVSAVELFVVEKKLFIHEYISGYVR 465
DB 435 VLIGIFLFGQQTQGVGNWINGAIFLNTMTFQNSPATITVFTLPEVFKRETSRLYR 494
QY 466 VSSYFLGKLLSLLPMRLMPSIIFTCIVFMGLGPKADAFVPMFTLMWVAYSASSMAL 525
DB 495 CDYFLGKTIAE-LFLFLVVPFLTAIAYPLIGLRPGVDHFTALALVTLVANVSTSGY 553
QY 526 AIAAGSVSVATLMTICFVFMVIFSGLLVNLTTIASWLSLQYFIPRYGFTALQHNE 585
DB 554 LISCACSSSTSMALSGPPVILPFLFGGFFLNSGSGVPVYFKWLSYLSWFRYANEGLLIQ 613
QY 586 FLGQNFQCPG-LNATGNPCNYATCTGEEVLVKQGLDLSWGLWKKHVALACMIVFLTIA 644
DB 614 W--ADVKGPEITCTLSN-----TTCPSGSEVILETLNFSASDLFPDFIGLALLIVGRISA 667

QY 645 YLKL 648
DB 668 VIAL 671

RESULT 6

WHIT_LUCCU STANDARD; PRT; 677 AA.
AC Q05360;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE White protein.
GN W.
OS Lucilia cuprina (Greenbottle fly) (Australian sheep blowfly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Oestroidea;
OC Calliphoridae; Lucilia.
OX NCBI_TaxID=7375;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=97087158; PubMed=8933176;
RX Garcia R.L., Perkins H.D., Howells A.J.;
RT "The structure, sequence and developmental pattern of expression of
RT the white gene in the blowfly *Lucilia cuprina*.";
RL Insect Mol. Biol. 5:251-260(1996).
RN [2]
RP SEQUENCE OF 490-584 FROM N.A.
RX MEDLINE=90264941; PubMed=1971656;
RA Elizur A., Vacek A.T., Howells A.J.;
RT "Cloning and characterization of the white and topaz eye color genes
RT from the sheep blowfly *Lucilia cuprina*.";
RL J. Mol. Evol. 30:347-358(1990).
CC -!- FUNCTION: MAY BE PART OF A MEMBRANE-SPANNING PERMEASE SYSTEM
CC NECESSARY FOR THE TRANSPORT OF PIGMENT PRECURSORS INTO PIGMENT
CC CELLS RESPONSIBLE FOR EYE COLOR.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY, MDR SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch)

EMBL; U38899; AAA82057.1; -
EMBL; X53265; CAA37365.1; -
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR005284; Pigment_permease.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR TIGRfams; TIGR00955; 3a01204; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
KW Pigment; ATP-binding; Transmembrane; Transport.
FT NP_BIND 119 126 ATP (POTENTIAL).
FT TRANSMEM 431 451 POTENTIAL.
FT TRANSMEM 456 476 POTENTIAL.
FT TRANSMEM 506 526 POTENTIAL.
FT TRANSMEM 534 554 POTENTIAL.
FT TRANSMEM 563 583 POTENTIAL.
FT TRANSMEM 647 667 POTENTIAL.
SQ SEQUENCE 677 AA; 75365 MW; D16FC11C97EED51D CRC64;
Query Match 22.3%; Score 748.5; DB 1; Length 677;
Best Local Similarity 31.2%; Pred. No. 1.2e-43;
Matches 193; Conservative 124; Mismatches 226; Indels 75; Gaps 18;
QY 5 NVEVPIPVSGQNTNGFPATVSNLDKAFTEGAVLSFHNIYRVK---LKSGLPCRKPVEK 61

```
Db 63 NUDVFGVHQPSN-----WKQLVNRVKGVCNERHIP--KP-RK 99
Qy 62 EILSNINGIMKPG-LNAIILGPTGGKSSLLDVLARK-----DPSGLSGDVLINGAPRP 114
Db 100 HLKKNVGVAYGELLAVMGSSGAGTKTLLNALAFSARGVQVSPSSVR--MANGHPVD 156
Qy 115 A-NFKNSGVYVDDVVMGTLTVRENLOFSAALRLATTWTHHEKNERINRVIEELGLDKV 173
Db 157 AKEMOARCAVQDDDLFISGLTAREHLIFQATVRPRTMTQKQLQKRVQDQVQLSLIKC 216
Qy 174 ADSKVGTO-FIRGVSGEKRKTSIGMELTDPSSILSLDEPTTGLDSTANAVALLIKMS 232
Db 217 QNTIIGVGRVRLSGEGRKRLAFASEALTDPFLICDEPTSGLDSPMAASVYVQLKLS 276
Qy 233 KOGRTIIFSIHOPRYSIFKFLDPSLTLLASGRMLFHPGAQALGYFESAGYHCEAYNNPAD 292
Db 277 ORGKTIVLTHQPSSELFELFKILLMAGRVAFGLTGPVEADVDFSGAQCPPTYNPAD 336
Qy 293 PFLDINGSTAVANREEDFKATEIIEPSKQDKPLIEKLAIEYVNSFYKETAELHQL 352
Db 337 FYVQVL-----AVVPGRE-----IESDRISKICDNFAVGKVSREMEQNFQKI 379
Qy 353 S-----GGEKKKITVFEKISYTTSCHQLRWVSKSFKULLGNPOASIAQIIVTVGLV 408
Db 380 AAKTDGLQKDDETTLTKASWTFQF-----RAIMRWSWISTLKEPLLVKRVLIQTMMVAVL 435
Qy 409 IGAIYFGLKNDGTGIONRAGVLPFLTTNOCFSSVSAV-ELFVVEKKLFTHIEYISYRVS 467
Db 436 IGLIFLNPQMTQGVWNINGALFELTNMTFQNVFVAVINVTSELPVFMRETRSLYRCD 495
Qy 468 SYFLGKLSDLLPMRLPSSIIFTCIVYFMLGLKPKADAFVMMFTLMVMYVAYSASSMALAI 527
Db 496 TYFLGKLTAE-LPLFLVFPFLFIAIAYPMIGLRPGITHELSALAVTLVANVSTSGVLI 554
Qy 528 AGQSVSVATLL---MTICFVPMIFSGLLVNTIISWLSWLOVFSIPRYGFTALQHN 584
Db 555 SCASTSTSNALSGVPLTIPTF---LLFGGVFLNSGVVPVYFKWLSYFVSFRYANEGLLIN 611
Qy 585 EFL-----GONFCPLGNAT 598
Db 612 QNADVQPGELTCTNTT 629
```

RESULT 7

```
WHIT ANOGA
ID WHIT ANOGA STANDARD; PRT; 695 AA.
AC Q27256; Q17006;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE White protein.
GN w.
OS Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.
OX NCBI_TaxId=71165;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Suakoko / G3;
RX MEDLINE=96423158; PubMed=8825759;
RA Besansky N.J., Bedell J.A., Benedict M.Q., Mukabayire O., Hilfiker D.,
RA Collins F.H.;
RT "Cloning and characterization of the white gene from Anopheles
RT gambiae";
RL Insect Mol. Biol. 4:217-231(1995).
CC -!- FUNCTION: MAY BE PART OF A MEMBRANE-SPANNING PERMEASE SYSTEM
CC NECESSARY FOR THE TRANSPORT OF PIGMENT PRECURSORS INTO PIGMENT
CC CELLS RESPONSIBLE FOR EYE COLOR.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.eib-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

```
DR EMBL; U29486; AAC46995.1; -.
DR EMBL; U29485; AAC46994.1; -.
DR EMBL; U29484; AAC47423.1; -.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR005284; Pigment_permease.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR TIGRFAMs; TIGR00955; 3a01204; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
DR KW Pigment; ATP-binding; Transmembrane; Transport.
FT NP_BIND 133 140 ATP (POTENTIAL).
FT NP_BIND 288 295 ATP (POTENTIAL).
FT TRANSMEM 444 464 POTENTIAL.
FT TRANSMEM 474 494 POTENTIAL.
FT TRANSMEM 524 544 POTENTIAL.
FT TRANSMEM 552 572 POTENTIAL.
FT TRANSMEM 581 601 POTENTIAL.
FT TRANSMEM 669 689 POTENTIAL.
FT CARBOHYD 472 472 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 100 100 N -> S (IN REF. 1; AAC47423).
FT CONFLICT 691 693 SRS -> YAR (IN REF. 1; AAC47423).
SQ SEQUENCE 695 AA; 77218 MW; EB8B9517239B2961 CRC64;
```

Query Match 22.2%; Score 745; DB 1; Length 695;

Best Local Similarity 30.0%; Pred. No. 2.2e-43;
Matches 201; Conservative 130; Mismatches 265; Indels 74; Gaps 19;

```
Qy 6 VEVPIPVSGNTNGFPATVSNDLKAFTEGAVLSFHNICYRVKLKSGFLPCKRKPVEKILS 65
Db 78 IDVF---GEAPTQKP-----REPLCTLRNCCTR--QRKDFNP-----RKHLK 117
Qy 66 NINGIMKPG-LNAIILGPTGGKSSLLDVLARKDPS-GLSGDVL--INGAPRPA-NFKCN 120
Db 118 NVTGVAKSGELLAVMGSSGAGTKTLLNALAFSPGPKISPNVAVRALNGVPVNAEQLRAR 177
Qy 121 SGYVQDDVVMGTLTVRENLOFSAALRLATTWTHHEKNERINRVIEELGLDKVADSKVGT 180
Db 178 CAYVQDDLFIPSLTTREHLLFQAMLRMRGRDVPASVKQHRVQEVQLQELSLVKCADTIIGA 237
Qy 181 Q-FIRGVSGEGRKRTSIGMELITDPSILSLDEPTTGLDSTANAVALLIKMSKQRTII 239
Db 238 PGRKIGLSGGERKRLAFASELTIDPHLLCDEPTSGLDSPMAHSLVQLVKGMAMKGTII 297
Qy 240 FSIHQPRYSIFKFLDPSLTLLASGRMLFHPGAQALGYFESAGYHCEAYNNPADPFLDIN 299
Db 298 LTIHQPSSELYCFDKILLVAEGRVAFGLSPYQAEFPFSQLGIPCPNPNPADFPVQML- 356
Qy 300 GDSTAVANREEDFKATEIIEPSKQD--KPLIEKLAIEYVNSFYKETAELHQLSGSEK 357
Db 357 -----AIAPAKEACRDMIKKICDSFAVSPITAREVLETASVAGKMD 398
Qy 358 K-----KKITVPKEISYTFCHQLRWVSKSFKULLGNPOASIAQIIVTVGLVIGAIY 413
Db 399 EPYMLQVEGVSGTGYRSSWWTQPCILWRWSLVSKDPMVLRVLLQATNAVATLIGSIY 458
Qy 414 FGLKNDSTGIONRAGVLPFLTTNOCFSSVSAV-ELFVVEKKLFTHIEYISYRVSYPFG 472
Db 459 FGQVLDQGVWNINGSLFLFLTNMTFQNVFVAVINVTSELPVFMREKRSRLYRVDYFLG 518
Qy 473 KLLSDLLPMRLPSSIIFTCIVYFMLGLKPKADAFVMMFTLMVMYVAYSASSMALATAAGQS 532
Db 519 KTIAB-LPLFAVFPVFTSITPYMIGLRTGATHYLTTLFIVTLVANVSTSGYLISCASS 577
```



```

RL Gene 185:77-85(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=DBA/2;
RX MEDLINE=96359154; PubMed=87031120;
RA Savary S., Denizot F., Luciani M.-F., Mattei M.-G., Chimini G.;
RT "Molecular cloning of a mammalian ABC transporter homologous to
RL Drosophila white gene.";
RM Mamm. Genome 7:673-676(1996).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=21092576; PubMed=11162488;
RA Lorkowski S., Rust S., Engel T., Jung E., Tegelkamp K., Galinski E.A.,
RA Asemann G., Cullen P.;
RT "Genomic sequence and structure of the human ABCG1 (ABC8) gene.";
RL Biochem. Biophys. Res. Commun. 280:121-131(2001).
RN [4]
RP INDUCTION, AND PROBABLE FUNCTION.
RX MEDLINE=20261604; PubMed=10799558;
RA Venkateswaran A., Repa J.J., Lobbaccaro J.-M.A., Bronson A.,
RA Mangelsdorf D.J., Edwards P.A.;
RT "Human white/murine ABC8 mRNA levels are highly induced in
RT lipid-loaded macrophages. A transcriptional role for specific
RT oxysterols.";
RL J. Biol. Chem. 275:14700-14707(2000).
RN [5]
RP REVIEW.
RX MEDLINE=21474438; PubMed=11590207;
RA Schmitz G., Langmann T., Helmerl S.;
RT "Role of ABCG1 and other ABC family members in lipid metabolism.";
RL J. Lipid Res. 42:1513-1520(2001).
CC -/- FUNCTION: Transporter involved in macrophage lipid homeostasis. Is
CC an active component of the macrophage lipid export complex. Could
CC also be involved in intracellular lipid transport processes. The
CC role in cellular lipid homeostasis may not be limited to
CC macrophages.
CC -/- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -/- TISSUE SPECIFICITY: EXPRESSED MAINLY IN BRAIN, THYMUS, LUNG,
CC ADRENALS, SPLEEN AND PLACENTA. LITTLE OR NO EXPRESSION IN LIVER,
CC KIDNEY, HEART, MUSCLE OR TESTES.
CC -/- INDUCTION: Strongly induced in macrophage cell line RAW264.7
CC during cholesterol influx. Induction is mediated by the liver X
CC receptor/retinoid X receptor (LXR/RXR) pathway.
CC -/- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCG (WHITE)
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U34920; AAB47738.1; -.
DR EMBL; Z48745; CAAB8636.1; -.
DR EMBL; AF233659; AAK27442.1; -.
DR MGI; 107704; Abcg1.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR005284; Pigment_permease.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_tran; 1.
DR SMART; SM00382; AAA; 1.
DR TIGRfam; TIGR00955; 3a01204; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
KW ATP-binding; Transmembrane; Transporter.
FT DOMAIN 1 414 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 415 433 POTENTIAL.
FT DOMAIN 434 444 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 445 465 POTENTIAL.
FT DOMAIN 466 494 CYTOPLASMIC (POTENTIAL).

```

```

FT TRANSMEM 495 513 POTENTIAL.
FT DOMAIN 514 521 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 522 543 POTENTIAL.
FT DOMAIN 544 555 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 556 574 POTENTIAL.
FT DOMAIN 575 637 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 638 657 POTENTIAL.
FT DOMAIN 658 666 CYTOPLASMIC (POTENTIAL).
FT NP BIND 118 125 ATP (POTENTIAL).
SQ SEQUENCE 666 AA; 74033 MW; EDDC6AFBD43950B6 CRC64;

Query Match 21.3%; Score 712.5; DB 1; Length 666;
Best Local Similarity 28.9%; Pred. No. 3.5e-41;
Matches 198; Conservative 146; Mismatches 243; Indels 97; Gaps 24;

QY 3 SSVNEVFPVSGQNTNGFPATVSDNLKAPTE-----GAV-LSFHNIYRYVK---- 47
DB 38 SSVND---EVETDLLNGHLKKVDNN---FTAQRPSSLPRAAVNIEFADLSYSVEGPW 91
QY 48 -LKSGLPCRPKEKEILSNINGIMKPG-LNAILGPTGGKSSLDVLAARKDPGLSGD 105
DB 92 WKKGKY-----KTLKLGISGKFNSELGELVAIMPSPGAGKSTLMNLAGVRE-TGMKA 142
QY 106 VLINGAPRAN-FKCNISGVVDDVVMGTLTVRENLFQSAALRLATTNTNHEKRIIRV 164
DB 143 VLINGMPRLRCFRKVCYIMQDDMLPLHLTVQEAAMVSAHLKQE--KDEGRREMVKEI 200
QY 165 IEEGLDKVADSKVGTQFIRGVSGGERKRTSIGELITDPSITLSLDEPTTGLDSTANAV 224
DB 201 LTALGLLPCANRTGS-----LSGQQRRLALALELVNPPVWFDEPTSGLDASCFQV 255
QY 225 LLLKRMKQKQRTIIFSIHQPRYSIFKLFDLSITLASRLMPHGPAPQALGYFESAGYHC 284
DB 256 VSLMKGLAGGRSIVCTTHQSAKLFELFDQLYVLSQGVYRGKSVNLVPLRDLGLNC 315
QY 285 EAYNPADFFLDIINGD-----STAVANRE-----EDFKATEIIEPSPKQDKPLRK 331
DB 316 PTHNPADFVMEVASGEYGDQNSRLVRAVRGMCADYKRDGGDTVNPFLMHRPAED 375
QY 332 LAEIVVNSFYKETKAEHLQHLGGKSKKKTIVFKEISVTTGFCHELRVWVSKSPKLNLCN 391
DB 376 SASMEGCHSF-----SASCLTQFC-----ILFKKTFISIMRD 407
QY 392 PQASIAQIVTVVLGLVIGAIYFGLKNDSTGIONRAGVLFFLTNNQCFSSVSAVEL-FVV 450
DB 408 SVLTHLRITSHIGILLGLLYLGIGNEAKKVLNSGFLFSGMLFLMFAALMPTVLTPL 467
QY 451 EKKLPIHYISGYRVSSYFGLKLSLDLPLMRMLPSIIFTCIIVYFMLGLKPKADAFVWM 510
DB 468 EMSVFLREHLNMYWYSLKAYLAKTMAD-VPFQIMPVAYCSIVYWMTS-QPSDAYRVFLP 525
QY 511 FTL-MMVAYSASSMALATAAGOSVVSVALTMTICFVPMIFPSGLLVNLTITIASWLSLQ 569
DB 526 AALGTMTSILVQAQSLGILLGAASTSLQVATFVGPVTAIPVLLFSGFVFDITIPALQWMS 585
QY 570 YFSIPRYGFTALQHNFLQNGFCPCGLNATGNPCNYA-TC-TGEEYLVKQGDIDSPMGL 626
DB 586 YISVYRGFEGVILSIY-----GLDREDLH-CDIAETCFQKSEAILRE-LDVENAKL 636
QY 627 WKNHVALACMIVFIITAIYKLKLP 650
DB 637 VLDFIVLGLGIFISLRILAIYFVLRY 660

RESULT 10
ABGI HUMAN
ID ABGI HUMAN STANDARD; PRT; 678 AA.
AC P45844; Q9BXK6; Q9BXK7; Q9BXK8; Q9BXK9; Q9BXL0; Q9BXL1; Q9BXL2;
AC Q9BXL3; Q9BXL4;
DT 01-NOV-1995 (Rel. 32, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE ATP-binding cassette, sub-family G, member 1 (White protein homolog)

```

DE (ATP-binding cassette transporter 8).
GN ABCG1 OR ABC8 OR WH1.
OS Homo sapiens (Human).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE OF 3-678 FROM N.A. (ISOFORMS 1 AND 4).
RC TISSUE=Retina;
RX MEDLINE=96256850; PubMed=9659545;
RA Chen H.M., Rossier C., Laiotti M.D., Lynn A., Chakravarti A.,
RA Perrin G., Antonarakis S.E.;
RT "Cloning of the cDNA for a human homologue of the Drosophila white
RT gene and mapping to chromosome 21q22.3";
RL Am. J. Hum. Genet. 59:66-75(1996).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=20289799; PubMed=10830953;
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Groner Y.,
RA Soeda E., Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K.,
RA Polley A., Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA Reichwald K., Rump A., Schillhabel M., Schudy K., Asakawa S.,
RA Rosenthal A., Kudoh J., Shibuya K., Mitsuyama S., Antonarakis S.E.,
RA Shintani A., Sasaki T., Nagamine K., Kwasaki K., Asakawa S.,
RA Minooshima S., Shimizu N., Nordstiek G., Hornischer K., Brandt P.,
RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,
RA Ramser J., Beck A., Klages S., Hennig S., Riesemann L., Daggand E.,
RA Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA Lehrach H., Reinhardt R., Yaspo M.-L.;
RT "The DNA sequence of human chromosome 21";
RL Nature 405:311-319(2000).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=20408883; PubMed=10950923;
RA Berry A., Scott H.S., Kudoh J., Tallor I., Korostishevsky M.,
RA Wattenhofer M., Guipponi M., Barras C., Rossier C., Shibuya K.,
RA Wang J., Kawasaki K., Asakawa S., Minooshima S., Shimizu N.,
RA Antonarakis S.E., Bonne-Tamir B.;
RT "Refined localization of autosomal recessive nonsyndromic deafness
RT DPNB10 locus using 34 novel microsatellite markers, genomic
RT structure, and exclusion of six known genes in the region";
RL Genomics 68:22-29(2000).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=21192304; PubMed=11279031;
RA Porsch-Oezcuermmez M., Langmann T., Heimerl S., Borsukova H.,
RA Kaminski W.E., Brobnik W., Honer C., Schumacher C., Schmitz G.;
RT "The zinc finger protein 202 (ZNF202) is a transcriptional repressor
RT of ATP binding cassette transporter A1 (ABCA1) and ABCG1 gene
RT expression and a modulator of cellular lipid efflux";
RL J. Biol. Chem. 276:12427-12433(2001).
RN [5]
RP SEQUENCE FROM N.A. (ISOFORMS 2; 3; 4; 5; 6 AND 7).
RX MEDLINE=21092576; PubMed=11162488;
RA Lorkowski S., Rust S., Engel T., Jung E., Tegelkamp K., Galinski E.A.,
RA Assmann G., Cullen P.;
RT "Genomic sequence and structure of the human ABCG1 (ABCG8) gene";
RL Biochem. Biophys. Res. Commun. 280:121-131(2001).
RN [6]
RP SEQUENCE OF 33-678 FROM N.A.
RC TISSUE=Fetal brain;
RX MEDLINE=97186700; PubMed=9034316;
RA Croop J.M., Tiller G.E., Fletcher J.A., Lux M.L., Raab E.,
RA Goldenson D., Arciniegas S., Son D., Wu R.;
RT "Isolation and characterization of a mammalian homologue of the
RT Drosophila white gene";
RL Gene 185:77-85(1997).
RN [7]
RP INDUCTION, AND PROBABLE FUNCTION.
RX MEDLINE=20261604; PubMed=10799558;
RA Venkateswaran A., Repa J.J., Lobaccaro J.-M.A., Bronson A.,
RA Mangelsdorf D.J., Edwards P.A.;
RT "Human white/murine ABC8 mRNA levels are highly induced in
RT lipid-loaded macrophages. A transcriptional role for specific
RT oxysterols";
RL J. Biol. Chem. 275:14700-14707(2000).
RN [8]
RP INDUCTION, AND PROBABLE FUNCTION.
RX MEDLINE=20105556; PubMed=10639163;
RA Klucken J., Buechler C., Orso E., Kaminski W.E.,
RA Porsch-Oezcuermmez M., Liebisch G., Kapinsky M., Diederich W.,
RA Brobnik W., Dean M., Allikmets R., Schmitz G.;
RT "ABCG1 (ABCG8), the human homologue of the Drosophila white gene, is a
RT regulator of macrophage cholesterol and phospholipid transport";
RL Proc. Natl. Acad. Sci. U.S.A. 97:817-822(2000).
RN [9]
RP REVIEW.
RX MEDLINE=21474438; PubMed=11590207;
RA Schmitz G., Langmann T., Heimerl S.;
RT "Role of ABCG1 and other ABCG family members in lipid metabolism";
RL J. Lipid Res. 42:1513-1520(2001).
RN [10]
RP FUNCTION: Transporter involved in macrophage lipid homeostasis. Is
CC also be involved in intracellular lipid transport processes. The
CC role in cellular lipid homeostasis may not be limited to
CC macrophages.
CC -! SUBUNIT: May form heterodimers with several heterologous partners
CC of the ABCG subfamily.
CC -! SUBCELLULAR LOCATION: Integral membrane protein. Predominantly
CC localized in the intracellular compartments mainly associated with
CC the endoplasmic reticulum (ER) and Golgi membranes.
CC -! ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=7;
CC Comment=Additional isoforms seem to exist;
CC Name=1;
CC IsoID=P45844-1; Sequence=Displayed;
CC Name=2; Synonyms=J;
CC IsoID=P45844-2; Sequence=VSP_000047, VSP_000051;
CC Name=3; Synonyms=ABDE;
CC IsoID=P45844-3; Sequence=VSP_000048, VSP_000051;
CC Name=4; Synonyms=G;
CC IsoID=P45844-4; Sequence=VSP_000051;
CC Name=5; Synonyms=F;
CC IsoID=P45844-5; Sequence=VSP_000049, VSP_000051;
CC Name=6; Synonyms=HI;
CC IsoID=P45844-6; Sequence=VSP_000046, VSP_000051;
CC Name=7; Synonyms=C;
CC IsoID=P45844-7; Sequence=VSP_000050, VSP_000051;
CC -! TISSUE SPECIFICITY: EXPRESSED IN SEVERAL TISSUES.
CC -! INDUCTION: Strongly induced in monocyte-derived macrophages during
CC cholesterol influx. Conversely, mRNA and protein expression are
CC suppressed by lipid efflux. Induction is mediated by the liver X
CC receptor/retinoid X receptor (LXR/RXR) pathway.
CC -! SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCG (WHITE)
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X91249; CAA62631.1; ALT_INIT.
CC EMBL; AP001746; BAA95530.1; ALT_INIT.
CC EMBL; AB038161; BAB13728.2; ALT_INIT.
CC EMBL; AJ289137; CAC00730.1; ALT_INIT.
CC EMBL; AJ289138; CAC00730.1; JOINED.
CC EMBL; AJ289139; CAC00730.1; JOINED.
CC EMBL; AJ289140; CAC00730.1; JOINED.
CC EMBL; AJ289141; CAC00730.1; JOINED.
CC EMBL; AJ289142; CAC00730.1; JOINED.
CC EMBL; AJ289143; CAC00730.1; JOINED.
CC EMBL; AJ289144; CAC00730.1; JOINED.


```
DR Genew; HGNC:113884; ABCG4.
DR InterPro; IPR003593; AAA ATPase.
DR InterPro; IPR003439; ABC_transporter.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
KW ATP-binding; Glycoprotein; Transmembrane; Transport.
FT DOMAIN 1 393 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 394 414 1 (POTENTIAL).
FT DOMAIN 415 425 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 426 446 2 (POTENTIAL).
FT DOMAIN 447 472 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 473 493 3 (POTENTIAL).
FT DOMAIN 494 503 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 504 524 4 (POTENTIAL).
FT DOMAIN 525 532 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 533 553 5 (POTENTIAL).
FT DOMAIN 554 617 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 618 638 6 (POTENTIAL).
FT DOMAIN 639 646 CYTOPLASMIC (POTENTIAL).
FT NP_BIND 102 109 ATP (POTENTIAL).
FT CARBOHYD 422 422 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 646 AA; 71895 MW; 9CCBC6E150772611 CRC64;

Query Match 21.0%; Score 702; DB 1; Length 646;
Best Local Similarity 27.5%; Pred. No. 1.8e-40;
Matches 175; Conservative 142; Mismatches 247; Indels 72; Gaps 16;

QY 37 LSFHNICYRVKLSGFLPC-RKPVEKEILSNING-IMKPLNAILGPTGGKSSLLDLVA 94
DB 61 IFEVLSYV--REG--PCWRKRGYKTLKLSGFCRELIGMPGSGAGSTFWNLA 116
QY 95 ARKDPGSLGVDLINGAPRP-ANFKNSGVYVDDVMTGLTVRENLOPSAALRLATTMT 153
DB 117 GYRE-SGMKQILVNGRPRLTRFMSCYIMQDDMLLPHLTAVLEAMVMYANLKLSE--K 173
QY 154 NHEKNERINRVIEELGLDKVADSKVGTQPIRGVSGGERKRTSIGMELITDPSILSDEPT 213
DB 174 QBVKELVEITLALGLMSCSHRTAL-----LSGQKRRLAIALELVNPNPVMFDEPT 228
QY 214 TGLDSTANAVLLLRMSKQRTTIFSIHQPRYSIFKLFDSLTLLASRLMHFGPAQEA 273
DB 229 SGLDSASCFQVSLMKSLAQGRTTICTHQPSAKLFENFDKLYLSQGCIFKGVVNL 288
QY 274 LQYFESAGYHCAYNNPADFFLDIINGD-----STAVALNREEDFKATEII 319
DB 289 IPLYKGLGLHCPYHNPADFIIEVASGEYDLENPMLFRAVQNGLCMAEKKSSPEKNEVP 348
QY 320 ESKQDKPLLEKLAETIYVNSPFYKETAELHQLSGEKKKTVPEKISYTTSPFCHQLRW 379
DB 349 APCPPCPPEVDPI-----ESH-----TFATSTLTQFCI 376
QY 380 VSKRGFKNLLGNPQASIAQIIVTVLGVIGAIYFGLKNDSTGIONRAGLVPLTTNQCF 439
DB 377 LFKRFLSLTRDLTVLTHLRFMHSVHVGVLGLYLHLHGDDASKVFNNTGCLFSLMLFMF 436
QY 440 SSVSAVEL-FVVEKKLFIHEYISGYRVSSYFLGKLLSLLPMRLPSIIFTCIYVFMGL 498
DB 437 AALMPTVLTLPLEMAVFMEHNLNYSKAYYLAKTMAD-VFFQVVCVPVYCSIVVMTG 495
QY 499 LKPKADAFVMMFTLMWYASASSMALATAAQSVSVVATLMTTCFVPMIFPSGLLNL 558
DB 496 QPAETSRFLFGSALATATATVALVAQSLGLLGAASNLQVATFPGVPTAIPVLVPSGFFVSF 555
QY 559 TTIAGSLMWLFPSIPRYGTALQHNELFGQFCPLGNATGNPNQYATCT-GEYLVKQ 617
DB 556 KTIPTYLQSSVLSYRVYGFEGVILTIY-----GME-RGDLTCLERCPRPQSIIR 607
QY 618 GIDLSPWGLKQNHVALACMIVFLTATYKLLFLKK 653
DB 608 ALDVEDAKLYMDFLVLGIFFLALRLAYLVLYRVK 643
```

RESULT 12

```
ABGS_MOUSE
ID ABGS_MOUSE STANDARD; PRT; 652 AA.
AC Q99PE8;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE ATP-binding cassette, sub-family G, member 5 (Sterolin-1).
GN ABCG5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Liver;
RX MEDLINE=20578753; PubMed=11138003;
RA Lee M.-H., Lu K., Hazard S., Yu H., Shulenin S., Hidaka H., Kojima H.,
RA Allikmets R., Sakuma N., Pegoraro R., Srivastava A.K., Salen G.,
RA Dean M., Patel S.B.;
RT Identification of a gene, ABCG5, important in the regulation of
RT dietary cholesterol absorption."
RL Nat. Genet. 27:79-83(2001).
RN [2]
TISSUE SPECIFICITY, AND INDUCTION.
RX MEDLINE=20553648; PubMed=11099417;
RA Berge K.E., Tian H., Graf G.A., Yu L., Grishin N.V., Schultz J.,
RA Kwtierovich P., Shan B., Barnes R., Hobbs H.H.;
RT "Accumulation of dietary cholesterol in sitosterolemia caused by
RT mutations in adjacent ABC transporters."
RL Science 290:1771-1775(2000).
CC -!- FUNCTION: Transporter that appears to play an indispensable role
CC in the selective transport of the dietary cholesterol in and out
CC of the enterocytes and in the selective sterol excretion by the
CC liver into bile.
CC -!- SUBUNIT: May form heterodimers with ABCG8 or be tightly coupled to
CC ABCG8 along a pathway regulating dietary-sterol absorption and
CC excretion (By similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -!- TISSUE SPECIFICITY: Expressed in the intestine and, at lower
CC level, in the liver.
CC -!- INDUCTION: Upregulated by cholesterol feeding. Possibly mediated
CC by the liver X receptor/retinoic X receptor (LXR/RXR) pathway.
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCG (WHITE)
SUBFAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AF312713; AAC53097.1; -.
MGD; MGI:1351659; Abcg5.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transporter.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
KW ATP-binding; Glycoprotein; Transmembrane; Transport.
FT DOMAIN 1 385 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 386 406 1 (POTENTIAL).
FT DOMAIN 407 422 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 423 443 2 (POTENTIAL).
FT DOMAIN 444 463 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 464 484 3 (POTENTIAL).
FT DOMAIN 485 504 EXTRACELLULAR (POTENTIAL).
```



```
FT TRANSMEM 505 525 4 (POTENTIAL).
FT DOMAIN 526 529 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 530 550 5 (POTENTIAL).
FT DOMAIN 551 622 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 623 643 6 (POTENTIAL).
FT DOMAIN 644 652 CYTOPLASMIC (POTENTIAL).
FT NP_BIND 87 94 ATP (POTENTIAL).
FT CARBOHYD 410 410 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 585 595 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 592 592 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 652 AA; 73244 MW; 80CE37ADCC19771E CRC64;

Query Match 20.6%; Score 690.5; DB 1; Length 652;
Best Local Similarity 29.0%; Pred. No. 1.1e-39;
Matches 181; Conservative 141; Mismatches 247; Indels 55; Gaps 16;

QY 12 VSQNTNGFPATVSNLKAFTGAVLSFNHICVRYKLSG-----FLPCRKPEKEILSNI 67
Db 25 LEQGSVTGEARHS-----LGVLVSVSYNVRGPPWNNIKSCQKWDROLKDV 73
QY 68 NGIMKPG-LNAILGPTGGKSSLLDLVLAARKDPG--LSGDVLINGAP-RPANFKCNSGYV 124
Db 74 SLYIESQIMCIGSSGSKTLLDAISGLRLRTGTLEGEVFNCGELRRDQFCFSV 133
QY 135 VQDDVVMGTLTVRENLOFSAALRLATMTNHEKNERINRIVBELGLDKVADSKVGTQFIR 184
Db 134 LQSDVFLSSLVRETLRYTAMALCRSSADP-YNNKVEAVMTLSLSHVADQMIGSYNFG 192
QY 185 GVSGERKRTSGMELITDPSILSDPTGLDSTANAVALLKXMSQGRITIFSIIHQ 244
Db 193 GLSSGERRRVSTAAQLQDPKVMMLDEPTTGJDCMTANQIVLLLAELARRDRIVITIHQ 252
QY 245 PRYSIFKLDSTLLASGRMFHFGPAQEAALGYFESAGYHCEAYNPNPADFFLDIINGDSTA 304
Db 253 PRSELFQHFQDKAILTYGELVFCGTPPEMLGFNNGYPCPSHPNPPDFYMDLTSVDIQ- 311
QY 305 VALNRE-BDFKATEIIEPSKQDKLIEKLAIEVYNSSFYKETKABELHQLSGGEKKKJTV 363
Db 312 -SRREIETIKRVQMLECAFKESDIYHKI-----LENIERARYLKT 351
QY 364 FKEISYTT-----SFCHQLRWVSKRFFKLLGNLGNPQASIAQIIVTVVLGIVGAIYGLKND 419
Db 332 LPTVPFKTKDPPMGFKGLVLRVTRNLRNKQAVIMRLVQNLIMGLFLFYLLRVQNN 411
QY 420 ST--GIQNRAGVFLFTTNCQFSS--VSARELVEWEKCLFIHEYISGYRVSVSYFLGKLS 476
Db 412 TLKGAVQDRVGLLYQLVATPYTGMLNANVLEPMLRAVSDQSDQGLYHKWMLLAYVL- 470
QY 477 DLLPMRLPSIIFTCTIVFMGLKPKADAFVYMMFTLM---MVAYSASMALAIAGOSV 533
Db 471 HVLPPFSVIATVIFSSVCYVTLGLYPEVARFGYFSAALLAPHLIGBFLTLVLIGIVQNPNI 530
QY 534 VSVATLLMTICFVPMFIMISGLLVNLTTSWLSWLOYFSIPRYGTALQHNFLQNF-C 592
Db 531 VNSIVALLUSIS--GLLIGSGFIRTOEWPILKILGYTFQKYCCIEILVWNEFYGLNFTC 588
QY 593 PGLNATGNPNPCNYATCTGEEYLVK 616
Db 589 GGSNTSMLNHPALCATQGVQFIEK 612
```

```
RESULT 13
ABG5_RAT
AC Q99PE7, STANDARD; PRT; 652 AA.
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE ATP-binding cassette, sub-family G, member 5 (Sterolin-1).
GN ABG5.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
```

```
NCBI_TaxID=10116;
[1]
SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Small intestine;
MEDLINE=20578753; PubMed=11138003;
Lee M.-H., Lu K., Hazard S., Yu H., Shulenin S., Hidaka H., Kojima H.,
Allikmets R., Sakuma N., Pegoraro R., Srivastava A.K., Salen G.,
Dean M., Patel S.B.;
"Identification of a gene, ABG5, important in the regulation of
dietary cholesterol absorption.";
Nat. Genet. 27:79-83(2001).
CC -!- FUNCTION: Transporter that appears to play an indispensable role
in the selective transport of the dietary cholesterol in and out
of the enterocytes and in the selective sterol excretion by the
liver into bile..
CC -!- SUBUNIT: May form heterodimers with ABG8 or be tightly coupled to
ABG8 along a pathway regulating dietary-sterol absorption and
excretion (By similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABG5 (WHITE)
SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.ebi.ac.uk/announcements
or send an email to license@ebi.ac.uk).
```

```
EMBL; AF312714; AAGS3098.2; -
DR InterPro; IPR003593; AAA ATPase.
DR InterPro; IPR003439; ABC transporter.
DR Pfam; PF00005; ABC tran; 1.
DR ProDom; PD000006; ABC transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC TRANSPORTER 1; 1.
DR PROSITE; PS00893; ABC TRANSPORTER 2; 1.
KW ATP-binding; Glycoprotein; Transmembrane; Transport.
FT DOMAIN 1 385 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 386 406 1 (POTENTIAL).
FT DOMAIN 407 422 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 423 443 2 (POTENTIAL).
FT DOMAIN 444 463 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 464 484 3 (POTENTIAL).
FT DOMAIN 485 504 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 505 525 4 (POTENTIAL).
FT DOMAIN 526 529 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 530 550 5 (POTENTIAL).
FT DOMAIN 551 624 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 625 645 6 (POTENTIAL).
FT DOMAIN 646 652 CYTOPLASMIC (POTENTIAL).
FT NP_BIND 87 94 ATP (POTENTIAL).
FT CARBOHYD 585 585 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 592 592 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 652 AA; 73342 MW; 4D42FE2BAB0DAD59 CRC64;

Query Match 20.3%; Score 678.5; DB 1; Length 652;
Best Local Similarity 28.9%; Pred. No. 7.2e-39;
Matches 185; Conservative 141; Mismatches 238; Indels 77; Gaps 22;

QY 2 SSSNVEVPIPVSGQNTNGFPATVSNLKAFTGAVLSFNHICVRYKLSG-----FLPCRK 57
Db 21 SQSLE-----EGSVTGEARHS-----LGVLVSVSYNVRGPPWNNIKSCQ 63
QY 58 PVEKEILSNINGIMKPGIN-AILGPTGGKSSLLDLVLAARKDPG--LSGDVLINGAP-RP 114
Db 64 KWRDKILKDVSLYIESGQTMCLGSSGSKTLLDAISGLRLRTGTLEGEVFNCGELRR 123
QY 115 ANFKCNSGYVQDDVVMGTLTVRENLOFSAALRLATMTNHEKNERINRIVTEELGDKVA 174
Db 124 DQFQDCVSYLQSDVFLSSLVRETLRYTAMALRRSSADP-YDKKVEAVLTELSSLSHA 182
```

175	Qy	DSKVGTQFTRGVSGGERKRTSIGMELITDPSILSDERTPTGLDSSSTANAVLILLKRMQSK	234
183	Db	QDMIGNYNPGGJSSGERRVSAIAQLQDPKVMMLDDEFTGLDQMTANHIVLLVLARR	242
235	Qy	GRTIIPSIHQPIYSIFKLPDLSLTLASGRLMFHGPAQEAALGYPESAGYHCBAVNNPADRF	294
243	Db	NRIVIVTHQPSRSELFPHFDKIALITYGELVFCGTPPEMLGFFNNCGYPCPEHSNPPDY	302
295	Qy	LDIINGDSTAVANRE-EDFKATEIIEPSKODKPLIEKLAEIYVNSFSYKETAELHQUS	353
303	Db	MDLTSVDTC--SRREIETIKRVQMLSEAFRQSDICHKI-----LE	341
354	Qy	GGEKKKIITVFEISYTTIS-----FCHOLRVWSKESFNKLLGNPOASIAQIIVTVVLGLV	408
342	Db	NIERTHKLTLPMVPFKTKNPPGMPFC-KLGVLLRRVTRNLMRNKQVVIMRLVQNLINGLF	400
409	Qy	IGAIYFGLKNDST---GIQNRAGVLFFLTNNOCFSS-VSAVELFVVEKKLFIHEYISGY	463
401	Db	L--IFVLLRVQNMLKGAVQDVRVGLLYQLVGATPYTGMNAVNLFPMLRAVSDQESDGL	458
464	Qy	YRVSSYFLGKLLSDLLPWRMLPSPITFCIVFMGLGAPKADAFVMMFTLM---WVAYSA	520
459	Db	YQKQWMLLAYVL-HALPFSIVATVIFSSVCYWTLLGYPEVAREGYFSAALLAPHLIGEFL	517
521	Qy	SSMALAIAGQSWSVATLLTTCIVFPMFIESGLLVNLTITIASWLSWLOVFSIPRYGFTA	580
518	Db	TLVLGWQONPNVNSIVALISIS--GLITSGFIRNIEBMPILPKILGYFTFOKVCCEI	575
581	Qy	LQHNEFLGNQF-CPGLN-ATGNPN-CNYA-----TCGT	610
576	Db	LVNVPFYGLNFTCCGSNTSVNPNMCSMTQIGQFIKTCPI	616

```

RESULT 14
YPC3 CAEEL
ID YPC3 CAEEL STANDARD; PRT; 598 AA.
AC Q11180;
DT 01-NOV-1997 (Rel. 35, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Putative ABC transporter C05D10.3 in chromosome III.
DE C05D10.3.
GN Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RN SEQUENCE FROM N.A.
RN RP STRAIN=Bristol N2;
RN RC
RA Du Z.;
RA RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
RA [2]
RN RL
RN REVISIONS.
RA Waterston R.;
RA Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC ENBL; U13645; AAA20989.2; -.
CC WormPep; C05D10.3; CE29170.
CC InterPro; IPR003593; AAA_ATPase.
CC InterPro; IPR003439; ABC_transporter.
CC InterPro; IPR005284; Pigment_permease.
CC Pfam; PF00005; ABC_tran; 1.
CC ProDom; PD000006; ABC_transporter; 1.
CC -----

```

NCBI_TaxID=9606;
 (1)
 SEQUENCE FROM N.A., AND VARIANT GLU-604.
 TISSUE=Liver;
 MEDLINE=20553648; PubMed=11099417;
 Berge K.E., Tian H., Graf G.A., Yu L., Grishin N.V., Schultz J.,
 Kwikovich P., Shan B., Barnes R., Hobbs H.H.;
 "Accumulation of dietary cholesterol in sitosterolemia caused by
 mutations in adjacent ABC transporters.";
 Science 290:1771-1775(2000).
 (2)
 SEQUENCE FROM N.A., VARIANTS SITOSTEROLEMIA HIS-389; HIS-419 AND
 PRO-419, AND VARIANT GLU-604.
 TISSUE=Liver;
 MEDLINE=20578753; PubMed=11138003;
 Lee M.-H., Lu K., Hazard S., Yu H., Shulenin S., Hidaka H., Kojima H.,
 Allikmets R., Sakuma N., Pegoraro R., Srivastava A.K., Salen G.,
 Dean M., Patel S.B.;
 "Identification of a gene, ABCG5, important in the regulation of
 dietary cholesterol absorption.";
 Nat. Genet. 27:79-83(2001).
 (3)
 REVIEW.
 MEDLINE=2147438; PubMed=11590207;
 Schmitz G., Langmann T., Heimerl S.;
 "Role of ABCG1 and other ABCG family members in lipid metabolism.";
 J. Lipid Res. 42:1513-1520(2001).
 (4)
 VARIANTS SITOSTEROLEMIA GLN-146; HIS-389; PRO-419; HIS-419 AND
 SER-550, AND VARIANT GLU-604.
 MEDLINE=21344600; PubMed=11452359;
 Lu K., Lee M.-H., Hazard S., Brooks-Wilson A., Hidaka H., Kojima H.,
 Ose L., Stalenhoef A.F.H., Miettinen T., Bjorkhem I., Bruckert E.,
 Pandya A., Brewer H.B. Jr., Salen G., Dean M., Srivastava A.K.,
 Patel S.B.;
 "Two genes that map to the STSL locus cause sitosterolemia: genomic
 structure and spectrum of mutations involving sterolin-1 and
 sterolin-2, encoded by ABCG5 and ABCG8, respectively.";
 Am. J. Hum. Genet. 69:278-290(2001).
 CC -1- FUNCTION: Transporter that appears to play an indispensable role
 in the selective transport of the dietary cholesterol in and out
 of the enterocytes and in the selective sterol excretion by the
 liver into bile.
 CC -1- SUBUNIT: May form heterodimers with ABCG8 or be tightly coupled to
 ABCG8 along a pathway regulating dietary-sterol absorption and
 excretion.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -1- TISSUE SPECIFICITY: Strongly expressed in the liver, lower levels
 in the small intestine and colon.
 CC -1- DISEASE: Defects in ABCG5 are a cause of sitosterolemia
 (MIM:210250); also known as phytosterolemia or shellfish
 sterolemia. It is a rare autosomal recessive disorder
 characterized by increased intestinal absorption of all sterols
 including cholesterol, plant and shellfish sterols, and decreased
 biliary excretion of dietary sterols into bile. Sitosterolemia
 patients have hypercholesterolemia, very high levels of plant
 sterols in the plasma, and frequently develop tendon and tuberos
 xanthomas, accelerated atherosclerosis and premature coronary
 artery disease.
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCG (WHITE)
 SUBFAMILY.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL; AF320293; AAC40003.1; --
 EMBL; AF312715; AAC53099.1; --
 Genew; HGNC:13886; ABCG5.

DR MIM; 605459; --
 DR MIM; 210250; --
 DR GO; GO:0030299; P:cholesterol absorption; NAS.
 DR InterPro; IPR003593; AAA_ATPase.
 DR InterPro; IPR003439; ABC_transporter.
 DR Pfam; PF00005; ABC_tran; 1.
 DR ProDom; PD000006; ABC_transporter; 1.
 DR SMART; SM00382; AAA; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER_1; FALSE_NEG.
 DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
 KW ATP-binding; Glycoprotein; Transmembrane; Transport; Polymorphism;
 KW Disease mutation.
 FT DOMAIN 1 383
 FT TRANSMEM 384 404
 FT DOMAIN 405 421
 FT TRANSMEM 422 442
 FT DOMAIN 443 462
 FT TRANSMEM 463 483
 FT DOMAIN 484 503
 FT TRANSMEM 504 524
 FT DOMAIN 525 528
 FT TRANSMEM 529 549
 FT DOMAIN 550 623
 FT TRANSMEM 624 644
 FT DOMAIN 645 651
 FT NP_BIND 86 93
 FT CARBOHYD 584 584
 FT CARBOHYD 591 591
 FT VARIANT 146 146
 FT VARIANT 389 389
 FT VARIANT 419 419
 FT VARIANT 419 419
 FT VARIANT 550 550
 FT VARIANT 604 604
 FT VARIANT 604 604
 SQ SEQUENCE 651 AA; 72503 MW; 950B8F0CB6A1536 CRC64;
 Query Match 20.1%; Score 674.5; DB 1; Length 651;
 Best Local Similarity 29.0%; Pred. No. 1.4e-38;
 Matches 181; Conservative 137; Mismatches 251; Indels 55; Gaps 18;
 QY 13 SQGWTNGPPATVNDLKAFTGCAVLSPFNICRYVKLSGFLPCRPVKEILSNINGIMK 72
 DB 21 SQSLGEGAPATAP---EPHSLGILHASYSVSHRPPWMDITSRQQTROILKQVSLYVE 77
 QY 73 PG-LNAILGPTGGKSSLLDLAARKDPSG-LSGDVLING-APRANFKNSGVVQDDV 129
 DB 78 SGQIMCILGSGSGKNTLLDMSGLGRAGTFLGEVYNGRALRRREQQDFSYVLQSDT 137
 QY 130 VNGTLTVRENFQSAALRLATTMTNHEKRNINRVIELGLDKVADSKVGTQFIRGVSGG 189
 DB 138 LLSSLTVRETLHYTALLAIRGNPG-SFOKKVEAVMAELSLSHVADRLLIGNVSLGGISTG 196
 QY 190 ERKRTSIGMELITDPSILSLDEPTTGLDSTANAVLLLLKEMSKQGRITIIISHIQPRYSI 249
 DB 197 ERRRVISIAAQLQDPKVMFLFDEPTTGLDCMTANQIVLLVLLBELARRNRIVLITIHQPREL 256
 QY 250 FKLFDLTLASGLRMFHGPAQEAALGYPESAGYHCEAYNNPADFLDIINGDSTAVALNR 309
 DB 257 FQLFDKXIALSFGELIFCGTFAEMLDLPNDGCGYCPCHSNPFPDFMDLTSVDQTQ-----SK 312
 QY 310 EEDFKATEIIPSKQDKPLIEKLAIIVYNSFYKETAELHQLSGEGEKKXITVFKETSY 369
 DB 313 ERE-----IETSKR---VQMIESAYKKAICHKT-----LKNIERMKHLKTLPMVPF 356
 QY 370 TT-----SFCQLRWVSKSKSPKLLGNPOASTAQIIVTVLGLVIGAIYFGLKNDST---- 421
 DB 357 KTKDSPGVFSGKLVLLRRTNLRNKLAVITRLLQNLIIMGLFL--LFFVLIRVRSNVLKG 414

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 1, 2003, 18:45:26 ; Search time 43.2855 Seconds
(without alignments)
1455.232 Million cell updates/sec

Title: US-09-856-927-2
Perfect score: 3350
Sequence: 1 MSSSNVEFIPVSGQNTNGF.....MIVIFLTAYLKLFLKKYS 655

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues
Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 76:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2841.5	84.8	656	JC7860	brain multidrug re
2	835.5	24.9	1049	S19421	ATP-dependent perm
3	803	24.0	687	PFYFW	white protein - fr
4	795.5	23.7	737	T46101	ABC transporter-li
5	773	23.1	646	C86441	probable ABC trans
6	759	22.7	687	D96553	hypothetical prote
7	743	22.2	649	A84509	probable ABC trans
8	738	22.0	725	T47652	ABC transporter-li
9	734.5	21.9	739	T45891	ABC transporter-li
10	731	21.8	678	H96552	hypothetical prote
11	719.5	21.5	708	T47650	ABC transporter-li
12	715	21.3	635	T08934	hypothetical prote
13	712	21.3	740	T02567	probable ATP-bindi
14	712	21.3	755	G84791	probable ABC trans
15	710.5	21.2	609	E96742	probable ABC trans
16	706	21.1	638	G02068	white homolog - hu
17	702	21.0	646	JC7777	ATP binding cassete
18	700.5	20.9	720	T47648	ABC transporter-li
19	698.5	20.9	547	T31543	hypothetical prote
20	667	19.9	559	B88474	protein COSD10.3 [
21	667	19.9	725	C84423	probable ABC trans
22	657	19.6	659	E86313	hypothetical prote
23	655	19.6	608	T34391	hypothetical prote
24	653.5	19.5	1294	S77690	probable membrane
25	653	19.5	662	T47649	ABC transporter-li
26	650.5	19.4	590	B96573	protein F12M16.17
27	650.5	19.4	1450	T45888	ABC transporter-li
28	649.5	19.4	633	T19189	hypothetical prote
29	642.5	19.2	577	T04229	ABC-type transport

30	636.5	19.0	639	2	G88839	protein C10C6.5 [1
31	630.5	18.8	695	2	T21109	hypothetical prote
32	624	18.6	658	2	T31958	hypothetical prote
33	618.5	18.5	610	2	T19333	hypothetical prote
34	605.5	18.1	1501	2	S50992	SNQ2 protein - yea
35	603	18.0	1530	2	S52239	brefeldin A resist
36	596	17.8	1511	2	A53151	pleiotropic drug r
37	594	17.7	1530	2	T52010	hypothetical prote
38	588	17.6	1564	2	S55517	probable transport
39	583	17.4	1469	2	H96622	probable ABC trans
40	582	17.4	1443	2	T02491	probable ABC trans
41	579	17.3	1333	2	S63403	probable membrane
42	572.5	17.1	1420	2	T02644	ABC-type transport
43	569	17.0	1529	2	S69688	hypothetical prote
44	567	16.9	1413	2	G84790	probable ABC trans
45	559.5	16.7	1466	2	T30566	ATP-binding cassete

ALIGNMENTS

RESULT 1

JC7860
brain multidrug resistance protein, BMDP - pig
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 18-Nov-2002 #sequence_revision 18-Nov-2002 #text_change 31-Mar-2003
C:Accession: JC7860
R:Eisenblatter, T.; Galla, H.J.
Biochem. Biophys. Res. Commun. 293, 1273-1278, 2002
A:Title: A new multidrug resistance protein at the blood-brain barrier.
A:Reference number: JC7860; MUID:22050127; PMID:12054514
A:Accession: JC7860
A:Molecule type: mRNA
A:Residues: 1-656 <EIS>
A:CROSS-references: GB:AJ420927
A:Experimental source: brain
C:Comment: This protein, a new transport protein of the ATP-binding cassette (ABC) super
exclusion of xenobiotics from the brain and participates in drug transport across the bl
C:Genetics:
A:Gene: bmdp .

Query Match	84.8%	Score	2841.5	DB 2	Length	656			
Best Local Similarity	84.1%	Pred. No.	1.3e-191						
Matches	552	Conservative	44	Mismatches	59	Indels	1	Gaps	1
QY	1	MSSSNVEFIPVSGQNTNGFPATVNSDLKAFTEGAVLSFHNICVRVKLSGFLPCRKPVE	60						
DB	1	MSSSNVQVSIPIWKRNTNGLPGSSNELKTSAGGAVLSFHDICVRVKVSGFLPCRKTYE	60						
QY	61	KEILSNINGIMKPGNLAILGPTGGKSSLLDLAARKDPGSLGDLVINGAPRPANFKCN	120						
DB	61	KEILTNIINGIMKPGNLAILGPTGGKSSLLDLAARKDPHGLSGDLVINGAPRPANFKCN	120						
QY	121	SGYVQDDVNVGTLTVRENLFQSAALRLATMTNHEKNERINRVIEELGLDKVADSKVGT	180						
DB	121	SGYVQDDVNVGTLTVRENLFQSAALRLPTMTNHEKNERINRVIEELGLDKVADSKVGT	180						
QY	181	QFIRGVSGGERKRTSIGMELITDPSILSLDSEPTTGLDSDSTANAVLLILKEMSKQRTIIF	240						
DB	181	QFIRGVSGGERKRTSIGMELITDPSILFLDSEPTTGLDSDSTANAVLLILKEMSKQRTIIF	240						
QY	241	SIHQPRYSIFKLPDSTLTLLASGRMLFHPAQEALGFESAGYHCEAYNNPADFLDIING	300						
DB	241	SIHQPRYSIFKLPDSTLTLLASGRMLFHPGAREALGFASIGYNCPEYNNPADFLDVING	300						
QY	301	DSTAVALNR-BEDPKATIIIEPSKQDKPLIEKLAIIYVNSSPFYKETAELHQLSGGSKKK	359						
DB	301	DSSAVLSRADRDEGAQPEEPKEDTPLLIDKLAIFYNSSFPPKDTKVELDQFSGGRKKK	360						
QY	360	KITVFKIEISYTTSTFCHQLRWVSKRSFKNLLGNPQASIAQIIVTVVLGLVIGAIYFGLKND	419						
DB	361	KSSVYKEVYTTSTFCHQLRWISRRSFKNLLGNPQASVAQIIVTIIILGLVIGAIYFGLKND	420						

A:Experimental source: strain Canton S
R:O'Hare, K.

submitted to the EMBL Data Library, June 1985

A:Reference number: S10240

A:Accession: S10240

A:Molecule type: DNA

A:Residues: 1-24, 'LIFRIPYHCRVTAD', 30-687 <OH2>

A:Cross-references: EMBL:X02974; NID:g10873; PIDN:CAA26716.1; PID:g10874

A:Experimental source: strain Canton S

C:Genetics:

A:Gene: white; w

A:Cross-references: FlyBase:FBgn0003996

A:Introns: 24/3; 116/3; 334/3; 439/3; 483/3

C:Superfamily: fruit fly white protein; ATP-binding cassette homology

C:Keywords: ATP; glycoprotein; nucleotide binding; P-loop; transmembrane protein

F:113-317/Domain: ATP-binding cassette homology <ABC>

F:130-137/Region: nucleotide-binding motif A (P-loop)

F:261-265/Region: nucleotide-binding motif B

F:167,93,472,554,651/Binding site: carbohydrate (Aen) (covalent) #status predicted

Query Match 24.0%; Score 803; DB 1; Length 687;

Best Local Similarity 33.2%; Pred. No. 2e-48;

Matches 199; Conservative 127; Mismatches 233; Indels 40; Gaps 14;

QY 61 KEILSNINGIMKPG-LNALIGPTGGKSSLLDLAARKDPSGL-----SGDVLLINGAPRPA 115

Db 110 KHLXNVCVAYPGELLAVMGSSGAGKTTLLNALAFR-SPQIGVSPSGMRLNGOPVDA 168

QY 116 -NFKCNSGVVQDDVMGTLTVRENLOPSAALRLATTNHEKNERINRVIEELGLDKVA 174

Db 169 KEMQARCAVQDDLFISLTAREHLIFQAVMRPHLTQRVARVQVIOELSLSCQ 228

QY 175 DSKVGTO-FIRGVSGERKRTSIGMELITDPSILSDLEPTGLDSDSTANAVLLKRMKS 233

Db 229 HTIIGVGRVKLSGGERKRLAFASEALTDPPLLICDEPTSLDFTAHVVQVLUKSQ 288

QY 234 QGRTIIFSIHQPRYIFKFLDLSLLASGRMFHGPQAQALGYFESAGYHCEAYNNPADF 293

Db 289 KGTIVLTHQPSSELSFELFKILLMAEGRAVFLGTPSEAVDFSVVGAQCPTNYPADP 348

QY 294 FLDIINGSTAVALNEEDFKATEIIEPSKQDKPLIEKLAIEIYVNSSFYKTKAELHOLS 353

Db 349 YVQVL-----AVPQRETESR-----DRIAKICDNFAISKVAR-DMEQLL 387

QY 354 GGEKKKKTIVFKEISYV-TSPCHLRWVSKRSFKNLLGNPQASIAQIIVTVVLGVTGA 411

Db 398 ATKNEKPLEQEPENGYTKATWFMQFRAVLWRSWLSVLKEPLLVKVRLLIQTWAILGL 447

QY 412 IVFGLKNDSTGIONRAGVLFLLTTCNOFSSVSA-VELFVVEKKLFIHEYISGVSYF 470

Db 448 IFLGQQLTQGVNMINGAIFLFTNMTFQNVFATINVTSELPVFMREARSRLYCDTVF 507

QY 471 LKLLSDLLPMRLPSIIFTCIVYFMLGLKPKADAFFVMMFTLMVAVSASSMALAIAG 530

Db 508 LKGTIAB-LPLFLTVPLFTALAYPMIGRAGVLHFNCLALVTLVAVNSTSGFLISCA 566

QY 531 QSVWSVATLLMTICFVFMVIFSGLVNLTITIASWLSWLYQFSIPRYGFTALQHNFLQGN 590

Db 567 SSSTSMALSVGPPVLIPIFLFGGFFLNGSVVPVYLKWLISLWFRYANEGLLINGQADVE 626

QY 591 FCPG-LNATGNPNPCYATCTGEEYLVKQIDISPMGLWKNHVALACMIVPIITIAYLKL 648

Db 627 --PGBISCTSSN-----TTCPSGKGVILETLNFSAADLPDYVGLAILIVSPRLVAYAL 679

RESULT 4

T46101

ABC transporter-like protein - Arabidopsis thaliana

N:Alternate names: protein T25B15.80

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000

C:Accession: T46101

R:Alcaraz, J.P.; Ciabault, G.; Cottet, A.; Mache, R.; Mewes, H.W.; Lemcke, K.; Mayer, K.

submitted to the Protein Sequence Database, January 2000

A:Reference number: 223021

A:Accession: T46101

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-737 <ALC>

A:Cross-references: EMBL:ALJ32972

A:Experimental source: cultivar Columbia; BAC clone T25B15

C:Genetics:

A:Map position: 3

A:Introns: 122/1; 146/3; 225/2; 277/2; 339/3; 422/2; 535/1; 628/3; 664/3

A:Note: T25B15.80

Query Match

Best Local Similarity 23.7%; Score 795.5; DB 2; Length 737;

Matches 210; Conservative 134; Mismatches 229; Indels 101; Gaps 21;

QY 24 VSNDLKAPTEGAV-----LSPFNICYRVKLKSGFLPCRPVKEKILSNINGIMK 72

Db 121 ILEDIEAATSSVVKFQAEPTFPYILKFDITYKYVTTKG-----MTSSSEKSLNGISSAY 176

QY 73 PG-LNALIGPTGGKSSLLDLAARKDPSGLSGDVLLINGAPRANFKCNSGVVQDDVVM 131

Db 177 PGEALLMGPSSGSKTLLNALGGRFNQONTGGSVSYNDKYSKHLKTRIGFVTDVLF 236

QY 132 GTLVRENLOPSAALRLATTNHEKNERINRVIEELGLDKVADSKVGTQIRGVSGGER 191

Db 237 PHLTVKETLTYTALLRUPKTLTQEQEQRASVIOELGLERCQDTMIGGSFVRGVSGGER 296

QY 192 KRTSIGMELITDPSILSDLEPTGLDSDSTANAVLLKRMKQKQRTIIFSIHQPRYSIFK 251

Db 297 KRVCIGNEMTNPSSLILDEPTSSLDSTTALKIVQMLHCAKAGKTIIVTTHQPSRLFH 356

QY 252 LPDSITLLASGRMFHGPQAQALGYFESAGYHCEAYNNPADPFLDIINGDSTAVALNREE 311

Db 357 RFDKLWLSRSLLYFGKASEMSYFSSIGCSPLLAMNPABFLDLVNGNMNDIS----- 411

QY 312 DFKATEIIEPSKQDKPLIEKLAIEIYVNS-----SFYKTKAELHOLS----- 353

Db 412 -----VPSALKKMKIIRL-ELYVRNVKCDVETQYLEBAYKQIAYMKMKLMAPVPL 463

QY 354 GGEKKKKTIVFKE-----ISYTTSPFC-HQLRWVSKRS-----FKNLLGNPQASIAQIIVTVVLG 406

Db 464 DEEVKLMITCKREWGLSNWQYCLLSLRGKIKERRHDYFSLW-----RVTVQLSTAI-- 515

QY 407 LVIGAIYFGLKNDSTGIQ-NRAGVLFLLTTCNOFSSV-SAVELFVVEKKLFIHEYISGVY 464

Db 516 -ILGLLW--QSDITTSQRTSRSGLLFFIATVFWGFPFVFTAIPTFQERAMLSKERESNY 572

QY 465 RVSSYFLGKLLSDLLPMRLPSIIFTCIVYFMLGLKPKADAFFVMMFTLMVAVSASSMA 524

Db 573 RLSAYFVARTTSD-LPLDLILPVLFLVYVFMAGLRRAESFSLVTLVFLCIVAAQGLG 631

QY 525 LAIAAGOSVSVATLLMTICFVFMVIFSGLVNLTITIASWLSWLYQFSIPRYGFTAL---- 581

Db 632 LAIGASLMDLKATTLASVTVMTEMLAGGYFVK--KVPFFIAWIRFMFNHYTYLLVKV 689

QY 582 QHNFLGQFCPLGNATGNPCYATCTGEEYLVKQIDISPMGLWKNHVALACMIVLFL 641

Db 690 QYBEIM-----ESVNGEE-----IESGL-----KEVSALVAMLIIGYR 721

QY 642 TIAYLKLLFLKKYS 655

Db 722 LVAYFSLRRMKLHS 735

RESULT 5

C86441

probable ABC transporter [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C:Accession: C86441

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

RESULT 7
A:Accession: A84509
A:Title: ABC transporter [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C:Accession: A84509
R:Lin, X.; Kaul, S.; Rounley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; PMID:20083487; PMID:10617197
A:Accession: A84509
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-649 <STO>
A:Cross-references: GB:AE002093; NID:94558665; PIDN:AD22683.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g13610
A:Map position: 2
C:Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein F12L6.1; ATP-B

Query Match 22.2%; Score 743; DB 2; Length 649;
Best Local Similarity 33.5%; Pred. No. 3.1e-44;
Matches 203; Conservative 116; Mismatches 239; Indels 48; Gaps 18;
QY 61 KEILSNINGIMKP-GLNATLGTGGKSSLDVLAARKDPSGLSGDVLINGAP-RPANPK 118
DB 60 KHLVKGVTCTRAKPWILALVGSAGKSSLLILARLIPO--TGSVYVKNRPVDRANPK 117
QY 119 CNSGVYQDDVVMGTLTVRENIQFSAALATMTNHEKNERINRIVBELGLDKVADSK 178
DB 118 KISGYVTKQDTPLLTVEETLLFSKURL--KLPADELRSRVKSLVHGLGLEATATRV 175
QY 179 GTQFTRGVSGGRKRTSIGMELITDPSILSLDEPTGDSSTANAVLLLLKMSK-QQRT 237
DB 176 GDSVRGSGSGRRRYSIGVEVIHDPKVLILDEPTSGLDSTGALLIIDLKHMATRGRT 235
QY 238 IIFSTHOPRYSIFKLPDSLTLASGLMHPGAQAALGYFESAGVHCEAYNPPDFLDI 297
DB 236 ILLTHQGFRIYVKQFNSVILLANGSTLKQGSVDQLGVYLRNGLHPPLHENVIFAES 295
QY 298 INGDSTAVALNREEDFKATEIIEPSKQDKPLIEKLAIEYVNSFFYKETAELHQLSGGEB 357
DB 296 I--ESITKQRLQESRAAHVLP--QTTLQEKSEDSQGES--KSGKFTLQQLFQOTR 348
QY 358 KKKI--TVPKIEISYTSFCH-----QURWVSKRFKNLLGNPQASIAQIIVTVVLGVIGAI 412
DB 349 VADVGTMIATETFRDFANSRLTEETMILTRHPSKNIFRTKELFACRTVQMLGSGIVGLI 408
QY 413 YFGLKNDSTGIQNRAGVLPFLTNOCFSSVSASVELPVVEKLFIEHYISGYVRSYFLG 472
DB 409 FHLKDDLUKABERVGLFAFILTTLSTIEALPIFLOQERELMKETSGSYRVSAYVA 468
QY 473 KLLSDLLPMRLMPSIIFTCIVVFMGLGPKKADAFFVMFTLMWVAYSASSMALATAA--- 529
DB 469 NGLV-YLPFLLLILALFSTPVTVVLGNLPSFAPLHFLSLIILLYTANSVVVCFSAVP 527
QY 530 ----GQSVSVATLMTICFVPMFIFSGLLVNLTLTIASWLSWLYFSIPRYGFTALQHNE 585
DB 528 NPIVGNSVIG---VMGSPF----LFGSVFISHEIPGWIFPMHYISLFKYPFEGFLINE 580
QY 586 FLGQNFPCGLNATGNPCNYATCTGEYLVKQIDLSPMG---LWKHVALACMIVIFLT 642
DB 581 FSKSNK---LBYGFKC-----LVTEEDLLKE-----ERYGEESRRNVNIMLCFVLLYRP 629
QY 643 TAYLKL 648
DB 630 ISYVIL 635

RESULT 8
T47652
ABC transporter-like protein - Arabidopsis thaliana
N:Alternate names: protein T26112.10
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 19-May-2000
C:Accession: T47652
R:Monfort, A.; Casacuberta, E.; Puigdomenech, P.; Mewes, H.W.; Lemcke, K.; Mayer, K.F.X.;
submitted to the Protein Sequence Database, February 2000
A:Reference number: Z24471
A:Accession: T47652
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-725 <MON>
A:Cross-references: EMBL:AL132954
A:Experimental source: cultivar Columbia; BAC clone T26112
C:Genetics:
A:Map position: 3
A:Note: T26112.10
C:Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein F12L6.1; ATP-B

Query Match 22.0%; Score 738; DB 2; Length 725;
Best Local Similarity 29.1%; Pred. No. 8.1e-44;
Matches 192; Conservative 131; Mismatches 274; Indels 62; Gaps 13;
QY 36 VLSFHNICYRYKLXSGFLPCRPVKEILSNINGIMKFG-LNAIIGPTGGKSSLLDVLA 94
DB 72 VLFNNLYQDVTLRFRGFSRQGVKTLDDVSGEASDGLAVLGASGAGKSTLIDALA 131
QY 95 ARKDPGSLGSDVLINGAP--RPANPKNSGVYQDDVVMGTLTVRENIQFSAALRLATM 152
DB 132 GRVAGSLRGSVTLNGEKVLQSLRLKVISAYVMQDILLFPMLTVKETLMPFASBPRLPSL 191
QY 153 TNHEKNERINRIVBELGLDKVADSKVQFTRGVSGGRKRTSIGMELITDPSILSLDEP 212
DB 192 SKSKQKREVALIDQLGURNANVTIGDEHGRVSGGRRRRVSGIDIIHDPVILFDEP 251
QY 213 TTGLDSTANAVLLLLKMSKQGRITIIPIHQPRIYSIFKLPDSLTLASGLMHPGPAQE 272
DB 252 TSGLDSTNAPVWVQVLEIAQSGSIVIMS IHPSARIVELLDRLILSRGKSVFNGSPAS 311
QY 273 ALGFESAGYHCEAYNPPDFLDI-----NGDSTAVALNREEDFKATEIIEPS--- 322
DB 312 LPGFDFDFGRPIPEKENISEFALDLVRELGSEGTALVDNFNEKQKQNKISLIQSPOT 371
QY 323 ---KQDKPLIEKLAIEYVNSFFYKETAELHQLSGGKKKITVPKEISYTSFCHOLRW 379
DB 372 NKLDQDRSLSKEA---INASV---SRGK---VSGSRSNPTSMETVSYANPSLFFETFI 423
QY 380 VSKRSFKNLLGNPQASIAQIIVTVVLGVIGAIYFGLKNDSTGIQNRAGVLPFLTNOCF 439
DB 424 LAKRYMKNWIRMPDELVGTRITATVMTGCLLATVYMKLDHTPRGAQERLTLPAFVVPMPY 483
QY 440 SSVSAVELPVVEKLFIEHYISGYVRSYFGLKLLSLLPMRLMPSIIFTCIVVFMGL 499
DB 484 CCLDNVPVFIQERYIFLRETHNAYRTSSYVISHSLVS-LPQLLAPSLVFAISAITFTVGL 542
QY 500 KPKADAPFVMMFTLMWVAYSASSMALATAACOSVVSVATLMTICFVPMFIFSGLLVNL 559
DB 543 SGLEGFVYCLLIYASFWSGSSVTFISGVVPMIMLCYMSIITVLAICLLSLSGFYVARD 602
QY 560 TIASWLSWLYFSIPRYGFTALQHNEFLQGNFC-----PGLNATG----- 599
DB 603 RIFPVTWTFHYSILKYPIEAVLINEPDDPSRCFVRGVQVFDSTLLGGVSDGSKVKLET 662
QY 600 ----NPNCVATC--TGEEYLVKQGI--DLSPWGLWKNHVALACMIVIFLTIAYLKL 650
DB 663 LSKSLRTKITESTCLRTGSDLLAQOQTLSKWD-----CLWITFASGLFPRILF 712

RESULT 9
T45891
ABC transporter-like protein - Arabidopsis thaliana

RESULT 11

T47650
ABC transporter-like protein - Arabidopsis thaliana
N:Alternate names: protein T15C9.110
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 19-May-2000
C:Accession: T47650
R:Mewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.
submitted to the Protein Sequence Database, April 2000
A:Reference number: Z24470
A:Accession: T47650
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-708 <NEW>
A:Cross-references: EMBL:AL132970
A:Experimental source: cultivar Columbia; BAC clone T15C9
C:Genetics:
A:Map position: 3
A:Note: T15C9.110
C:Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein F12L6.1; ATP-B

Query Match 21.5%; Score 719.5; DB 2; Length 708;
Best Local Similarity 29.9%; Pred. No. 1.6e-42;
Matches 197; Conservative 124; Mismatches 282; Indels 55; Gaps 15;
QY 36 VLSFNICYRVKLSGFLPCRKPVE--KEILSNINGIMKPG-LNALIGPTGGKSLLDV 92
DB 62 LLSFNLSYVNVLRREFDSRRKTASVKTLLDITGEARDGEILAVLGGAGKSTLIDA 121
QY 93 LAARKDPSGLSDVLINGAP--RPNFNKNSGVVDDVVMGTLTVRENLOFSAAALRLAT 150
DB 122 LAGRAVEDSLKGTVTVLNGSKVLQSRLLKVISAVMDDLLFPMILTVEASEPRLPR 181
QY 151 TWTNHEKRIINRVIEELGLDKVADSKVCTQFIRGVSGGERKRTSGIMELITDPSLSLD 210
DB 182 SLPKSMKRVETLIDQLGRNAUTVIGDEHGVSGGERKRVISGIDIIHDPILLFLD 241
QY 211 EPTTGLDSTANAVILLKRMKSKQRTIIFSHQPRYSIFKLFDLSLTLLASGRMLFHPGA 270
DB 242 EPTSGLDSTNAFVWQVLRQAQSGSVIMSHQPSARIIGLLDRLIILSHGKSVENGSP 301
QY 271 QEALGVFESAGVHCEAYNPAPDFLDII-----NGDSTAVALNREEDFKATEIIEPSKOD 325
DB 302 VSLPSFSGFRPIPEKENITBFDLVIRELGSSSEGRDLVEFNEKMQQNOTARATTQS 361
QY 326 KPLIEKLAIIYVNSSFYKETKAEHLQSGEKKKTIIVFKEISYTTSPCHQLRWYKSRFP 385
DB 362 RVSLKEAIAASV-----SRGKL--VSGSGANPISMETVSSYANPPLAETFLAKRYI 412
QY 386 KNLLGNPQASIAQIIIVTVVLGIVGAIYFGLKNDSTGIQNRAGVLFLLTTCQFSSVASV 445
DB 413 KNWIRTPELIGRIGVTVMVGTLLATVYVRLONTPRGAQERWGFAGFSTWYCCADNI 472
QY 446 ELFPVVEKKLFIHEYISGYRVSSYFGLKLSDLLPMRMLPSIIFTCTIVYFMLGKPKADA 505
DB 473 PVFIQERYIFLRETHNARTSSYVISHALVS-LPOLLALSIFAFAATFTWTVGLSGGLES 531
QY 506 FPMFMFTLMWVAYSASSMALATAAGOSVSVATLLMTICFV-FMMIFSGLLVNLITIASW 564
DB 532 FPYCYLIIIAAAMSGSSVITFI-SGLIPNMVMSYMTIAYLSYCLLGGFYINRDIRPLY 590
QY 565 LSWLOYFSIPRYGFTALOHNEFLGQNF-----PGLNATG 599
DB 591 WLFPHVILSKYPYEAIVLINEEDDPSRCFVKGVQVDFGTLLAEVSHVMVKLLDITLSGL 650
QY 600 NNPCNVATC--TGEEYLVKQGI-DLSPWG-LMKNHVALACMIVFIPTIAYLKLFLKK 653
DB 651 GTKITESTCLRTGPDLLMQQGITQLSKWDCLW---ITLAWGL-FPRLFYLSILLFGSK 704

RESULT 12

T08934

hypothetical protein F27G19.20 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 17-Mar-2000
C:Accession: T08934
R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancroft, submitted to the Protein Sequence Database, May 1999
A:Reference number: Z16519
A:Accession: T08934
A:Molecule type: DNA
A:Residues: 1-635 <BEV>
A:Cross-references: EMBL:AL078467; GSPDB:GN00062; ATSP:F27G19.20
A:Experimental source: cultivar Columbia; BAC clone F27G19
C:Genetics:
A:Gene: ATSP:F27G19.20
A:Map position: 4
A:Introns: 38/3; 253/1; 304/1; 414/3
C:Superfamily: fruit fly white protein; ATP-binding cassette homology

Query Match 21.3%; Score 715; DB 2; Length 635;
Best Local Similarity 30.0%; Pred. No. 2.8e-42;
Matches 202; Conservative 125; Mismatches 274; Indels 72; Gaps 20;
QY 5 NVEVPVPSQGNWN-GPATVSNDLKAPTEGAVLSFHNICYRVKLSG---FLPCRKPVE 60
DB 7 SMDVETPIAKTNDSDRLSPFSI---FKANNPVTLKFNLYVTVKLKDSQSGCGKNDKTEE 63
QY 61 KEILSNINGIMKPG-LNALIGPTGGKSLLDVLAAR--KOPSGLSGDVLINGAPRPANF 117
DB 64 RTILKGLTGVKPGEILAMLGPGSGKTSLLTALGGRVGEKGKLTGNISYNNKPLSKAV 123
QY 118 KNSGVVQDDVVMGTLTVRENLOFSAAALRLATTVTNHEKNERINRVIEELGLDKVADSK 177
DB 124 KRTTFVQDDAALYNLVTETLVPTALLRPNSEPKQEKIKQAKAVMTGLDRCCKDTI 183
QY 178 VGTQPIRGVSGGERKRTSGIMELITDPSILSDEPTTGLDSTANAVILLKRMKSKQRT 237
DB 184 IGGPPLRGVSGGERKRVISGIEILINPSLLFLDEPTSGLDSTTAQRIIVSILMELARGRT 243
QY 238 IIFSHQPRYSIFKLFDLSLTLLASGRMLFHPGAQALGVFESAGVHCEAYN-NPADFFLD 296
DB 244 VVTHIQP-----SKGNPVYFGLGSNANDYFASVGSYPLVERINPDSFLD 289
QY 297 IINGDSTAVALNREEDFKATEIIEPSKODKPLIEKLAIIYVNSSPYK-----ETKA 347
DB 290 IANGKPLLV-----ISCWPSVGSDESQRPKAAALV---AFYKTNLLDSVINEVKG 338
QY 348 ELHQLSGGEKKKTIIVFKEISYTTSPCHQLRWYKSRFPKULLGNPQA--STAQIIVTVVL 405
DB 339 QDDLCKNKPRESSRVATNTYGDWPTTWQQFCVLLKRLKQRRHDSFGMKVQAQIF---IV 395
QY 406 GLVIGAIYFGLKNDSTGIQNRAGVLFLLTTCQFSSV-SAVELFVVEKKLFIHEYISGY 464
DB 396 SFLCGLLWQTK--ISRLODQIGLLFPFISSWAFPLPQQIPTFPQERAMLOKERSGNY 453
QY 465 RVSSYFGLKLSDLLPMRMLPSIIFTCTIVYFMLGKPKADAFPMFMFTLMWVAYSASSMA 524
DB 454 RLSPYFLSRVVD-LPMELILPTCLVITYWMAGLNHNLANPFTVLLVLLVHVLVSGGLG 512
QY 525 LAIAAGOSVSVATLLMTICFV-FMMIFSGLLVNLITIASWLSWLOYFSIPRYGFTALQHN 584
DB 533 LALGALVMDQKSATTLGSIIVMLTFLLAGGYVYVQHPV--FISWIKVYSGIYYTYKLL--- 567
QY 585 EPLGNQFCFGLNATGNN--PCNVATCTGEEYLV-KQGDILSPWGLMKNHVALACMIVFI 640
DB 568 -ILGQYTANELYPCDNGKLRCHVGDPEGIKHIGFNSGL-----VSALALTAMLVY 618
QY 641 LTIAYLKLFLKK 653
DB 619 RVIAIYALTRICK 631

RESULT 13

T02567

probable ATP-binding cassette protein T16B24.1 - Arabidopsis thaliana
N:Alternate names: protein F12L6.1
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 02-Mar-2001
R:Accession: T02567; T00545; C84816
R:Rounsley, S.D.; Lin, X.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, A.;
submitted to the EMBL Data Library, August 1998
A:Description: Arabidopsis thaliana chromosome II BAC T16B24 genomic sequence.
A:Reference number: Z14679
A:Accession: T02567
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-740 <R0U>
A:Cross-references: EMBL:AC004697; NID:g33402671; PIDN:AAC28975.1; PID:g33402672
R:Rounsley, S.D.; Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
submitted to the EMBL Data Library, July 1998
A:Description: Arabidopsis thaliana chromosome II BAC F12L6 genomic sequence.
A:Reference number: Z14168
A:Accession: T00545
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-362 <R0U>
A:Cross-references: EMBL:AC004218; NID:g3355463; PIDN:AAC27826.1; PID:g3355464
A:Experimental source: cultivar Columbia
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
Eisen, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: AB4420; MUID:20083487; PMID:10617197
A:Accession: C84816
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-740 <STO>
A:Cross-references: GB:AE002093; NID:g33402672; PIDN:AAC28975.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2G39350; T16B24.1; F12L6.1
A:Map position: 2
A:Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein F12L6.1; ATP-B
C:Keywords: ATP
F:110-310/Domain: ATP-binding cassette homology <ABC>

Query Match 21.3%; Score 712; DB 1; Length 740;
Best Local Similarity 27.4%; Pred. No. 5.6e-42;
Matches 193; Conservative 125; Mismatches 294; Indels 92; Gaps 17;
QY 12 VSQGNTPATVSNDLKAFTEGAVLSFNHCYRVKLGK-----SGFLPCR----- 56
DB 51 LQNDYNDGYMTVP-----FVLSFDNLTYNVSVRPKLDFRNLFRRTEDPEIAQT 101
QY 57 -KPEKEILSNINGIMKPG-LNAILGPTGGKSSLLDVLAARKDPSGLSDVLINGAPRP 114
DB 102 ARPRTKTLNNTSGETRDGEIMAVLGASGSKSTLIDALANRIAKSLGKTVKLGKLTQ 161
QY 115 AN-FKNCNSYVQDDVVMGTLTVRENLOFSALRLATTMTNEKNERINRVEELGDKV 173
DB 162 SRMLKVISAYVQDDLLFPMLTVEETLMPAERFRLPSLPSKKLRLVQALIDQLGIRNA 221
QY 174 ADSKVGTOPIRGVSGEERKRTSIGMELITDPSILSDEPTTGDSSTANAVLLKRMKS 233
DB 222 AKTIIGDEGHRGSGGERRRVIGIDIIHDPILLFIDEPTSGDSTSAFVVMVKLRKIAQ 281
QY 234 QORTIIFSIHQPRYSIFKLFDSLTLLASGRMLFHGPAQALGYFESAGHYCBAYNNPDP 293
DB 282 SGIIVMSIHQPSHRVGLGDLRLIFLSRGHTVYSGSPASLPFRFFTEFGSPIPENNRTEF 341
QY 294 FLDIIL-----NGDSTAVANREEDFKATEIIEPSKDKPLIEKLAETIYVNSFYKETK 346
DB 342 ALDLIRELEGSAGTGLIEFNK-----KQEWKQSGNSQPPITPP-SSPYMLTLKEAIA 396
QY 347 AELHQ-----LSGGEK-----KKKTIIVFKESYTSFCHQLRWVSKRSFKNLLGNPQASIA 397

DB 397 ASISRGKLVSGESVAHGGAATNTTTTAVPAPANPMWIEIKTLTKSRSLMNSRQPELPGI 456
QY 398 QIIVTVGLVIGLAIYFGLKNDSTGICQNRAGVLFLLTNQCFSSVSASVELFVVEKLPFH 457
DB 457 RIASVVIITGFTLATVFWBLDNRSPKGVQBRGLGFFAFAMSTMYFTCADALFVFLQERYIFMR 516
QY 458 EVISGYRVSVFGLKSLDILLPMRLPSIIFTCTIVYFMLGLKPKADAFFVMMFTLMVVA 517
DB 517 ETAYNAYRRSSVLSHAIVS-FPSLIPLSVAFAATTYWAVGLDGLTGLLFCYLLILASP 575
QY 518 YSASSMALAIAGOSVSVATLLMTICFVMMIFSGLLVNLTTIASLSWLOQYFSPRYG 577
DB 576 WSGSFTVFLSGWPSVMLGVTIVVAILAYFLFSGFFINRRIIDYWIWPHYMSLVKYP 635
QY 578 FTALOHNEFLGNFC--POLNATGNP-----CNYATC--TG 610
DB 636 YEAVLQNEFSDATKCFVRGVQIFDNTPLGELPEVMKJLGTVSKSLGVTISSTTCLTGTG 695
QY 611 EBYLVKQG-IDLSPWGLWKNHVALACMIVIFLTIA---YLKLLF 650
DB 696 SDILRQGVVQLSKWN-----CLFITVAFGFFRILF 727

RESULT 14

GB4791
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C:Accession: G84791
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
Eisen, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: AB4420; MUID:20083487; PMID:10617197
A:Accession: G84791
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-755 <STO>
A:Cross-references: GB:AE002093; NID:g4056489; PIDN:AAC98055.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2G37360
A:Map position: 2
C:Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein F12L6.1; ATP-B
Query Match 21.3%; Score 712; DB 2; Length 755;
Best Local Similarity 28.2%; Pred. No. 5.8e-42;
Matches 201; Conservative 127; Mismatches 273; Indels 112; Gaps 20;
QY 2 SSSNVEVPIPVSGNTNGFPATVSNDLKAFTEGAVLSFNHCYRVKLGSGFLP---CRKP 58
DB 78 SFPNSWASAPASSISSSPF-----VLSFTDITYSVKIOKKFNPLACCRRS 122
QY 59 VE-----KEILSNINGIMKPG-LNAILGPTGGKSSLLDVLAARKDPSGLSDVLINGA 111
DB 123 GNDSSVNTKILLANGISGEAREGEMAVLGASGSKSTLIDALANRIAKDSLRSITLNGE 182
QY 112 PRPANF-KCNSGYVQDDVVMGTLTVRENLOFSALRLATTMTNEKNERINRVEELGL 170
DB 183 VLESSMQKVISAYVQDDLLFPMLTVEETLMPAERFRLPSLPSKKLRLVQALIDQLG 242
QY 171 DKVADSKVGTOPIRGVSGEERKRTSIGMELITDPSILSDEPTTGDSSTANAVLLKLR 230
DB 243 RSAAKTVIGDGHGVRGSGGERRRVIGIDIIHDPILLFIDEPTSGDSTSAFVVMVKLR 302
QY 231 MSKQRTIIFSIHQPRYSIFKLFDSLTLLASGRMLFHGPAQALGYFESAGHYCBAYNNP 290
DB 303 IAQSGSIVMSIHQPSYRIMGLLQILFLSKGNTVYSGSPHLPQFFSEFKHIPENENK 362
QY 291 ADPFLDIINGDSTAVANREEDFKATEIIEPSKDKPLIE-----KLAETIYV----- 338
DB 363 TEFALDLI-----RELEYS-----TEGKPLVEFHKQWRAQAPSYNNNNKRN 405

```
QY 339 ---SSPYKETKALHQ---LSGGEKKKKTIVFKBI-SYTTSPCHQLRWVSKRSFKNLLGN 391
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 406 TNVSLKEAITASIRGKLVSGATNNSSNLTPSFQTFANPEWIEIVIGKEAILNSRRQ 465
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 392 POASIAQIIVTVVLGIVGAIYFGLKNDSTGIQNRAGVLFELTNOCFSSVSAVELFVVE 451
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 466 PELLGNRLGAVMTGIIATMTNLNDSKGAQERGLGFFAFAMSTTFYTCABAIPIVFLQE 525
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 452 KKLFIHEYISGYRYRVSSYFGLKLLSDLLPMRLPSIIFTCIVYFMLGLKPKADAFVMMF 511
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 526 RVIFMRETAYNRRSSVYLSQSIIS-IPALIVLSASFAATTFWAVGLDGGANGFFVYF 584
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 512 TLMWVAYSASSALAAQSVSVATLMTTCFVPMIFSGLLVNLTTIASWLSWLOQYP 571
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 585 TILASFAGSSFTVLSGVIPIVNMGLGFTVVVAILAYFLLSGFFISDRDIPVYMLMEHYI 644
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 572 SIPRYGFTALQNEFLGQN---FCPLNATGNP-----CN 604
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 645 SLVKPYEGVLQNEP--QNPTCFARGVQLFNSPLGEPNDVKVNLKSMGVLGTNVT 702
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 605 YATC--TGBEYLVKQGI-DLSPWG-LWKNHVALACMIVIFLTIA---YLKLLF 650
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 703 AETCVTTGIDILKQGGITDKWNCIM-----ITVANGFFERVLV 742
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 15
E96742
probable ABC transporter F17M19.11 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
C:Accession: E96742
R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.;
ansen, N.P.; Hughes, B.; Huizarr, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maity, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: E96742
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-609 <STO>
A:Cross-references: GB:AE005173; NID:96978921; PIDN:NAF34313.1; GSPDB:GN00141
C:Genetics:
A:Gene: F17M19.11
A:Map position: 1
C:Superfamily: fruit fly white protein; ATP-binding cassette homology

Query Match 21.2%; Score 710.5; DB 2; Length 609;
Best Local Similarity 31.0%; Pred. No. 5.4e-42;
Matches 190; Conservative 125; Mismatches 256; Indels 41; Gaps 16;

QY 56 RKPVEKEILSNINGIMKPG-LNAILGPTGGKSSLLDLVLAARKDPSGLSGDVLINGAPRP 114
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 23 RSTEERTILSGVTGMSPGEFNVLGPGSGSKSTLLNAVAGRLHGSNLTKILINDGKIT 82
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 115 ANFKCNISGVVQDDVMGTLTVRENLOFSAALRLATTMTNHEKNERINRINRVEELGDKVA 174
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 83 KOTLKRTPVQAODLLYPHLTIVRETLVFAALLRPSLRTRDVKLRAAESVISELGLTKE 142
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 175 DSKVGTQIRGVSGGKERKTSIGMELITDPSILSLDEPTTGIDSSANAVLLLLKRMK- 233
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 143 NTWGNVTFIRGISGGERKRVSAHELLINPSLLVDEPTSGDLATAALRLVQTLAGAHG 202
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 234 QGRTIIFSIHQPRYGIKFLFDSILIASRLMFHGPQAQALGYFESAGYCEAYNNPADF 293
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 203 KGKTVVTSIHQSSRVFQMFQFDVLLISEGKCLFVGKGRDAMAYFESVGFSPAFMNPADF 262
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 294 FLDIING--DSTAVALNREDFKATEIISPSKQDKPLIEKLAEL----YVNSSFYKETKA 347
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
Db 263 LLDLANGVCQDTEREKPNVRQTLVAYDTLLAPQVKTCIEVSHPPQDNARFVKT--- 319
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 348 ELHQLSGGEEKKKITVPKEISYTTISFCHQL-RWUSKRSFKNLLGNPQASIAQIIVTVVLG 406
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 320 ---RVNGGITTCTA-----TWFSQLCILLHRLKERRHESF-----DLURIFOVVAAS 365
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 407 LVIGAIYFGLKNDSTGIQNRAGVLFELTTN-QCFSSVSAVELFVVEKKLFIHEYISGYR 465
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 366 ILCLGLMWN--HSDYRDVHDLGLLFFISIFNGVLPSPFNAVPTFPQERAIPTRRASGMT 423
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 466 VSSYFLGKLLSDLLPMRLPSIIFTCIVYFMLGLKPKADAFVMMFTLMWVAYSASSMAL 525
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 424 LSSYFMAHVLSGSLMELVLPASFLT-FTYWMVYLRPGIVPFLTLTSLVLLVYLASQGLGL 482
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 526 AIAAGQSVSVSATLMTTCFVPMIFSGLLVNLTTIASWLSWLOQYFSIPRYGF---TALQ 582
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 483 ALGAIMDAKASTIIVTMLAFVLTGGYYN--KPSGMVMKTVSTTFYCYRLLVAIQ 540
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 583 HNEFLGQNFCEP--GLNATGNPCNYATCTGBEYLVKQGIIDLSPMGLWKNHVALACMIVIF 640
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 541 YGS--GBEILRMGLGCDKSGKQASAAATSGACRFVEEEVI--GDVGMWTSVGVLPPLMFFGY 596
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 641 LTIAYLKLLFLK 652
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 597 RVLAYLALRRIK 608
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

Search completed: August 1, 2003, 18:51:47
Job time : 45.7855 secs

Best Local Similarity 30.2%; Pred. No. 0.0023;
Matches 45; Conservative 19; Mismatches 61; Indels 24; Gaps 5;
QY 3 LGAEAYTASSMALAIA---TGQSVSVATLMTIAFVFMMLFSGLLVNLRTIGPWLWL 58
DB 629 LEAVHTSISMFCIAAIFQTGVAAMTAGSFMLITFV---PAGFAIPTDMPQWLK 684
QY 59 QYFSPRYGFTALQWNEFLG---QBFPCGFNVTDNSTCVNSYAICTGNEYLINQIGIELSP 115
DB 685 FWNPNISYAEIGLSVNEFLAPRWQKQP-TNVT-----LGRITLESRGINDYD 731
QY 116 WGLWKNHVALACMIIFLTIAVLKLLFLX 144
DB 732 YMYVWSLSALLGLTIIFNTIPTLALSFLX 760
RESULT 13
T47649
ABC transporter-like protein - Arabidopsis thaliana
N:Alternate names: protein T15C9.100
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 26-May-2000
C:Accession: T47649
R:Mewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.
submitted to the Protein Sequence Database, April 2000
A:Reference number: Z24470
A:Accession: T47649
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-662 <NEW>
A:Cross-references: EMBL:AL132970
A:Experimental source: cultivar Columbia; BAC clone T15C9
C:Genetics:
A:Map position: 3
A:Note: T15C9.100
C:Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein F12L6.1; ATP-B
Query Match 14.9%; Score 114.5; DB 2; Length 662;
Best Local Similarity 26.3%; Pred. No. 0.0023;
Matches 49; Conservative 23; Mismatches 63; Indels 51; Gaps 9;
QY 2 GLGAEAYTASSMALAIAIQSVSVS-----VATLLMTIAFV-FMMLFSGLLVNLRTI 51
DB 482 GLAGFIYILMIIFASFWSGCSFVTFVSGVPIPNWMSYMTFGYLSYCLLFSGFYVNRDRI 541
QY 52 GPWLSWLOYFSIPRYGFTALQWNEFLGQEF--PCFNVTDN----- 90
DB 542 HLYWIWIHYSILKPYEAVLHNEFDPSRCFVRGNQVFDNIMGVSETTKAKLETWS 601
QY 91 -----STCVNSYAICTGNEYLINQIGIE-LSPWG-LWKNHVALACMIIFLTIAVLK 139
DB 602 GYLGMELTESTC-----LTTGSDLLKQHGIEQLDKWGCLM---VTILA-WGFFFRILFYFS 652
QY 140 LLFLPK 145
DB 653 LLLGSK 658
RESULT 14
T47648
ABC transporter-like protein - Arabidopsis thaliana
N:Alternate names: protein T15C9.80
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 19-May-2000
C:Accession: T47648
R:Mewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.
submitted to the Protein Sequence Database, April 2000
A:Reference number: Z24470
A:Accession: T47648
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-720 <NEW>
A:Cross-references: EMBL:AL132970

A:Experimental source: cultivar Columbia; BAC clone T15C9
C:Genetics:
A:Map position: 3
A:Note: T15C9.80
C:Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein F12L6.1; ATP-B
Query Match 14.9%; Score 114.5; DB 2; Length 720;
Best Local Similarity 23.1%; Pred. No. 0.0025;
Matches 40; Conservative 25; Mismatches 67; Indels 41; Gaps 5;
QY 5 AEATASSMALAIAIQSVSVSATLMTIAFVFMMLFSGLLVNLRTIGPWLWLOYPSIP 64
DB 553 ASFWSSGSFVTFLLSGVWPHVMLGYTVVAILAYFLFSGFFINRDRIPQYWIWFHYSLV 612
QY 65 RYGFALQWNEFLGQEF--PCFNVTDN-----TCVNSYAI 99
DB 613 KYPYEAIVLQNEFSDPTFCFVRGVQLFDNSPIGELTYGMKLLDSVSRISSTCL 672
QY 100 CTGNEYLINQIGI-ELSPWGLWKNHVALACMI-----IFLTIAVLKLLFLKK 145
DB 673 TTGADVLLKQGGVTLQSKWN-----CLLITVGGFGLFRILFYLCLLGSK 716
RESULT 15
T47652
ABC transporter-like protein - Arabidopsis thaliana
N:Alternate names: protein T26112.10
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 19-May-2000
C:Accession: T47652
R:Monfort, A.; Casacuberta, E.; Puigdomenech, P.; Mewes, H.W.; Lemcke, K.; Mayer, K.F.X.
submitted to the Protein Sequence Database, February 2000
A:Reference number: Z24471
A:Accession: T47652
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-725 <MON>
A:Cross-references: EMBL:AL132954
A:Experimental source: cultivar Columbia; BAC clone T26112
C:Genetics:
A:Map position: 3
A:Note: T26112.10
C:Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein F12L6.1; ATP-B
Query Match 14.9%; Score 114.5; DB 2; Length 725;
Best Local Similarity 24.8%; Pred. No. 0.0026;
Matches 41; Conservative 25; Mismatches 44; Indels 55; Gaps 7;
QY 19 TGQSVSVS-----VATLLMTIAFV-FMMLFSGLLVNLRTIGPWLWLOYFSIPRYG 68
DB 562 SGSSVWTFISGVWPNIMLCYVMSITYLAYCLLGSFYVNRDRIPPYMTWPHYISILKPY 621
QY 69 TALQWNEFLGQEF--PCFNVTDN-----STCVNSYA 98
DB 622 EAVLINEFDPSRCFVRGVQVFDSTLLGGVSDSGVKLLLETLSKSLRTKITESTCLR--- 678
QY 99 ICTGNEYLINQIGI-ELSPWGLWKNHVALACMIIFLTIAVLKLLF 142
DB 679 --TGSDDLAAQGITQLSKWD-----CLWITFASGLPFRILF 712
Search completed: August 1, 2003, 18:51:48
Job time : 11.2145 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 1, 2003, 18:45:06 ; Search time 102.089 Seconds
(without alignments)
1655.666 Million cell updates/sec

Title: US-09-856-927-2
Perfect score: 3350
Sequence: 1 MSSSNVEFIPVSGQNTNGF.....MIVFLTIAYLKLFKKYS 655

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : SPTREMBL_23.*
- 1: sp_Archea.*
 - 2: sp_Bacteria.*
 - 3: sp_Fungi.*
 - 4: sp_Human.*
 - 5: sp_Invertebrate.*
 - 6: sp_Mammal.*
 - 7: sp_Mhc.*
 - 8: sp_Organelle.*
 - 9: sp_Phage.*
 - 10: sp_Plant.*
 - 11: sp_Rodent.*
 - 12: sp_Virus.*
 - 13: sp_Vertebrate.*
 - 14: sp_Unclassified.*
 - 15: sp_Rv1r.*
 - 16: sp_Bacteriap.*
 - 17: sp_Archeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3337	99.6	655	4	Q96TA8
2	3333	99.5	655	4	Q81X16
3	3329	99.4	655	4	Q96LD6
4	2841.5	84.8	656	6	Q8MIB3
5	2754	82.2	657	11	Q9R004
6	1787.5	53.4	650	11	Q8BK15
7	848	25.3	801	5	Q8T691
8	803	24.0	751	10	Q93YS4
9	797	23.8	687	5	Q94960
10	796	23.8	687	5	Q9NH94
11	795.5	23.7	737	10	Q9FT51
12	784	23.4	679	5	Q9BH97
13	784	23.4	692	5	P91892
14	781.5	23.3	648	10	Q9C6W5
15	773	23.1	646	10	Q9C6R7
16	771.5	23.0	679	5	Q81S30

17	770.5	23.0	567	10	Q9FGL7
18	765	22.8	670	5	O77423
19	759	22.7	687	10	Q9C9K2
20	756.5	22.6	695	10	Q8LMQ5
21	754	22.5	672	10	Q9L182
22	751.5	22.4	798	5	Q8T689
23	747	22.3	703	10	Q8RXN0
24	743	22.2	649	10	Q9SIT6
25	741.5	22.1	691	10	Q8RW19
26	741.5	22.1	1328	5	Q9NGP5
27	740	22.1	669	5	Q8WRP2
28	738	22.0	725	10	Q9M3D6
29	734.5	21.9	739	10	Q9LFG8
30	733	21.9	669	5	Q8WRR1
31	731.5	21.8	1528	5	Q8T677
32	731	21.8	678	10	Q9C8J8
33	731	21.8	678	10	Q8GX48
34	728.5	21.7	662	10	Q949Y4
35	728.5	21.7	1520	5	Q8T687
36	726	21.7	626	5	Q8T684
37	724	21.6	832	5	Q9VQV4
38	724	21.6	832	5	Q9UAF0
39	720	21.5	832	5	Q8MS50
40	719.5	21.5	668	10	Q9ARU4
41	719.5	21.5	708	10	Q9M2V5
42	715	21.3	635	10	Q9SZR9
43	714.5	21.3	594	10	Q9LJC3
44	713.5	21.3	1509	5	Q8T688
45	713	21.3	638	5	Q8T685

ALIGNMENTS

RESULT 1

Q96TA8	PRELIMINARY;	PRT;	655 AA.
ID	Q96TA8		
AC	Q96TA8		
DT	01-DEC-2001	(TRENBLrel. 19, Created)	
DT	01-DEC-2001	(TRENBLrel. 19, Last sequence update)	
DT	01-MAR-2003	(TRENBLrel. 23, Last annotation update)	
DE	ATP-binding cassette superfamily G (White) member 2.		
GN	ABCG2.		
OS	Homo sapiens (Human)		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=21201983; PubMed=11306452;		
RA	Komatsu H., Kotani H., Hara Y., Nakagawa R., Matsumoto M.,		
RA	Arakawa H., Nishimura S.;		
RT	Identification of breast cancer resistant protein/mitoxanthrone		
RT	resistance/placenta-specific, ATP-binding cassette transporter as a		
RT	transporter of NB-506 and J-107088, topoisomerase I inhibitors with an		
RT	indolecarbazole structure.		
RL	Cancer Res. 61:2827-2832 (2001).		
DR	EMBL; AB051855; BAB4693.1.		
DR	InterPro; IPR003439; ABC_transporter.		
DR	InterPro; IPR006162; Ppantne_attach.		
DR	Pfam; PF00005; ABC_tran.1.		
DR	ProDom; PD000006; ABC_transporter; 1.		
DR	PROSITE; PS00012; PHOSPHOPANTHETHEINE; 1.		
KW	ATP-binding.		
SQ	SEQUENCE 655 AA; 72314 MW; ABAF66B96034C5A8 CRC64;		
Query Match	99.6%;	Score 3337;	DB 4; Length 655;
Best Local Similarity	99.5%;	Pred. No. 1.2e-219;	
Matches 652;	Conservative 1;	Mismatches 2;	Indels 0; Gaps 0;
QY	1 MSSSNVEFIPVSGQNTNGFPATVNDLKAFTEGAVLSFHNTCYRVKLSGFLPCRPV 60		
DB	1 MSSSNVEFIPVSGQNTNGFPATVNDLKAFTEGAVLSFHNTCYRVKLSGFLPCRPV 60		


```
QY 1 MSSNVVEFIPVSGQNTNGFPATVNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKPVE 60
DB 1 MSSNVVEFIPVSGQNTNGFPATVNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKPVE 60
QY 61 KEILSNINGIMKPGNALILGPTGGKSSLLDLAARKOPSGLSGDVLINGAPRPANFKCN 120
DB 61 KEILSNINGIMKPGNALILGPTGGKSSLLDLAARKOPSGLSGDVLINGAPRPANFKCN 120
QY 121 SGYVVQDDVMGTLTVRENLOFSAALRLATTMTNHEKNERINRVIEELGLDKVADSKVGT 180
DB 121 SGYVVQDDVMGTLTVRENLOFSAALRLATTMTNHEKNERINRVIEELGLDKVADSKVGT 180
QY 181 QFIRGVSGGERKRTSIGMELITDPSILSDPTTGLDSTANAVALLKRMKSKQRTTIIF 240
DB 181 QFIRGVSGGERKRTSIGMELITDPSILSDPTTGLDSTANAVALLKRMKSKQRTTIIF 240
QY 241 SHQPRYSIFKLPDSLTLASGRLMFHGPAQALGYFESAGYHCEAYNNPADFFLDIING 300
DB 241 SHQPRYSIFKLPDSLTLASGRLMFHGPAQALGYFESAGYHCEAYNNPADFFLDIING 300
QY 301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAIEYVNSSFYKETKAEHLQLSGGEKKK 360
DB 301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAIEYVNSSFYKETKAEHLQLSGGEKKK 360
QY 361 ITVFKIEISYTTSFCHQLRWVSKRSFKNLLGNPQASIAQIIVTVLGLVIGAIYFGLKND 420
DB 361 ITVFKIEISYTTSFCHQLRWVSKRSFKNLLGNPQASIAQIIVTVLGLVIGAIYFGLKND 420
QY 421 TGIONRAGVLPFLTTNQCFSSVSARELVVEKKLFTHIEYISGYRVSSYFGLKLLSDLL 480
DB 421 TGIONRAGVLPFLTTNQCFSSVSARELVVEKKLFTHIEYISGYRVSSYFGLKLLSDLL 480
QY 481 MRMLPSIIFTCTIVYFVFLGLKPKADAFVMMFTLMVAVSASSMALAIAAGQSVSVATLL 540
DB 481 MRMLPSIIFTCTIVYFVFLGLKPKADAFVMMFTLMVAVSASSMALAIAAGQSVSVATLL 540
QY 541 MTICFVFMFIFSGLLVNLTTIASWLSWLYFSIPRYGFTALQHNFFLQNFPCGLNATGN 600
DB 541 MTICFVFMFIFSGLLVNLTTIASWLSWLYFSIPRYGFTALQHNFFLQNFPCGLNATGN 600
QY 601 NPCNYATCTGEEYLVKQIDLSPLWGLKXHVACMIVIFLTIAVLLKLLFKKYS 655
DB 601 NPCNYATCTGEEYLVKQIDLSPLWGLKXHVACMIVIFLTIAVLLKLLFKKYS 655
```

```
RESULT 4
Q8MB13 PRELIMINARY; PRT; 656 AA.
AC Q8MB13;
DB Q8MB13;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Brain multidrug resistance protein.
GN BMDP.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22050127; PubMed=12054514;
RA Eisenblatter T., Galla H.J.;
RT "A new multidrug resistance protein at the blood-brain barrier.";
RL Biochem. Biophys. Res. Commun. 293:1273-1278(2002).
DR EMBL; AJ240927; CAD12785.1; ...
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR006162; Ppantc_attach.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; 1.
KW ATP-binding.
```

```
SQ SEQUENCE 656 AA; 72392 MW; 118ADD5B53D9D67P CRC64;
Query Match 84.8%; Score 2841.5; DB 6; Length 656;
Best Local Similarity 84.1%; Pred. No. 8.7e-186;
Matches 552; Conservative 44; Mismatches 59; Indels 1; Gaps 1;
QY 1 MSSNVVEFIPVSGQNTNGFPATVNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKPVE 60
DB 1 MSSNVVEFIPVSGQNTNGFPATVNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKPVE 60
QY 61 KEILSNINGIMKPGNALILGPTGGKSSLLDLAARKOPSGLSGDVLINGAPRPANFKCN 120
DB 61 KEILSNINGIMKPGNALILGPTGGKSSLLDLAARKOPSGLSGDVLINGAPRPANFKCN 120
QY 121 SGYVVQDDVMGTLTVRENLOFSAALRLATTMTNHEKNERINRVIEELGLDKVADSKVGT 180
DB 121 SGYVVQDDVMGTLTVRENLOFSAALRLATTMTNHEKNERINRVIEELGLDKVADSKVGT 180
QY 181 QFIRGVSGGERKRTSIGMELITDPSILSDPTTGLDSTANAVALLKRMKSKQRTTIIF 240
DB 181 QFIRGVSGGERKRTSIGMELITDPSILSDPTTGLDSTANAVALLKRMKSKQRTTIIF 240
QY 241 SHQPRYSIFKLPDSLTLASGRLMFHGPAQALGYFESAGYHCEAYNNPADFFLDIING 300
DB 241 SHQPRYSIFKLPDSLTLASGRLMFHGPAQALGYFESAGYHCEAYNNPADFFLDIING 300
QY 301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAIEYVNSSFYKETKAEHLQLSGGEKKK 359
DB 301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAIEYVNSSFYKETKAEHLQLSGGEKKK 359
QY 361 ITVFKIEISYTTSFCHQLRWVSKRSFKNLLGNPQASIAQIIVTVLGLVIGAIYFGLKND 419
DB 361 ITVFKIEISYTTSFCHQLRWVSKRSFKNLLGNPQASIAQIIVTVLGLVIGAIYFGLKND 419
QY 421 TGIONRAGVLPFLTTNQCFSSVSARELVVEKKLFTHIEYISGYRVSSYFGLKLLSDLL 479
DB 421 TGIONRAGVLPFLTTNQCFSSVSARELVVEKKLFTHIEYISGYRVSSYFGLKLLSDLL 479
QY 481 MRMLPSIIFTCTIVYFVFLGLKPKADAFVMMFTLMVAVSASSMALAIAAGQSVSVATL 539
DB 481 MRMLPSIIFTCTIVYFVFLGLKPKADAFVMMFTLMVAVSASSMALAIAAGQSVSVATL 539
QY 541 MTICFVFMFIFSGLLVNLTTIASWLSWLYFSIPRYGFTALQHNFFLQNFPCGLNATG 599
DB 541 MTICFVFMFIFSGLLVNLTTIASWLSWLYFSIPRYGFTALQHNFFLQNFPCGLNATG 599
QY 601 NPCNYATCTGEEYLVKQIDLSPLWGLKXHVACMIVIFLTIAVLLKLLFKKYS 655
DB 601 NPCNYATCTGEEYLVKQIDLSPLWGLKXHVACMIVIFLTIAVLLKLLFKKYS 655
```

```
RESULT 5
Q9R004 PRELIMINARY; PRT; 657 AA.
AC Q9R004;
DB Q9R004;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Breast cancer resistance protein 1.
GN ABCG2 OR BCRP1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=FVB; TISSUE=Liver;
RX MEDLINE=99413474; PubMed=10485464;
RA Allen J.D., Brinkhuis R.F., Wijnholds J., Schinkel A.H.;
RT "The mouse Bcrp1/Mxr/Abcp gene: amplification and overexpression in
RT cell lines selected for resistance to topotecan, mitoxantrone, or
RT doxorubicin.";
RL Cancer Res. 59:4237-4241(1999).
```


DT 01-JUN-2001 (TREMELrel. 17, Last sequence update)
 DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
 DE White protein.
 GN WHITE.
 OS Ceratitis capitata (Mediterranean fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Tephritoidea; Tephritidae; Ceratitis.
 OX NCBI_TaxID=7213;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21135685; PubMed=11238408;
 RA Gomulski L.M., Pitts R.J., Costa S., Saccone G., Torti C.,
 RA Polito L.C., Gasperi G., Malacrida A., Kafatos F.C., Zwiebel L.J.;
 RT "Genomic Organization and Characterization of the white Locus of the
 RT Mediterranean Fruitfly, Ceratitis capitata."
 RT Genetics 157:1245-1255 (2001).
 CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
 DR EMBL; AF318276; AAK21872.1; -.
 DR EMBL; AF318275; AAK21871.1; -.
 DR InterPro; IPR003593; AAA_ATPase.
 DR InterPro; IPR003439; ABC_transporter.
 DR InterPro; IPR005284; Pigment_permease.
 DR Pfam; PF00005; ABC_tran; 1.
 DR ProDom; PD000006; ABC transporter; 1.
 DR SMART; SM00382; AAA; 1.
 DR TIGRFAMS; TIGR00955; 3a01204; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
 KW ATP-binding; Transport.
 SQ SEQUENCE 679 AA; 75005 MW; 3AE5509D7FC9F673 CRC64;

 Query Match 23.4%; Score 784; DB 5; Length 679;
 Best Local Similarity 31.4%; Pred. No. 3.8e-45;
 Matches 213; Conservative 138; Mismatches 251; Indels 76; Gaps 22;

 QY 11 PVSQNTN-GFP---ATVSNLKAFTT-----GAV-----LSFHNICYRVK-- 47
 DB 30 PYEQSSINQFSGKNGYGLTSPSPALTDNLITYSWYNLDVFGAVHQPQSSWKQLVNRKGV 89
 QY 48 -LKSGLPCPKVEKEILSNINGIMKPG-LNAILGPTGGKSSLDVLAARK-----DP 99
 DB 90 FCNERHIPAPR---KHLKNVSGVAYPGELLAVMGSSGAGKTTLLNAIFRSGKGVQISP 146
 QY 100 SGLSGDVLINGAPRPA-NFKCNSGYVQDDVVMGTLTVRENLOFSAALRLATTMTNHEKN 158
 DB 147 STIR---MLNGHPVDAKENQARCAVQDDDLFIGSLTAREHLIFQAVMPEPHMTQKQV 203
 QY 159 ERINRVIEELGLDKVADSKVGTQ-FIRGVSGGERKRTSIGMELITDPSILSDEPTGLD 217
 DB 204 QRVDQVQIDLSGKQNTLIGVGRVKGSLGGERKRLAFASEALTDPPLLICDEPTGLD 263
 QY 218 STANAVALLLXKMSKQGTITFISHOPRYSIFKLFDLSLTLASGRMLFHGPAQALGVF 277
 DB 264 SPMAHSVQVQLKSLQKGTVLTTHQPSSELFELFDKILLMAEGRVAFGLTPGGEAVDFP 323
 QY 278 ESAGYHCEAYNNPADFFLDINGDSTAVALEEDFKATEIIEPSKQDKLEKLAELVY 337
 DB 324 SVIATCPTNYPADFYQVLT-----AVVPGREVS-----ERAKICD 363
 QY 338 NSSFYKETKAEHLQSGGKKKTIYF-KE-----ISYTFSCHLRWKSRKFNLLGNP 392
 DB 364 NFAVGKVS-EMEQ-NFQKLVKSGFGKEDNGVYKASWPMQFRAVLRSWLSVLKEP 420
 QY 393 QASIAQIIVTVVLGVLGNIYGLKNDSTGIONRAGVLFLLTNQCFSSVSA-VELFVVE 451
 DB 421 LVLVRLQLTTWAVLIGLIFGQQLTQGVNNGINGAIFLFTNMTFQNAFATITVFTTE 480
 QY 452 KKLFIHEYISGYRVSSVPLGKLLDPLMRMLPSIIFCIYVFMGLKPKADAFVMMF 511
 DB 481 LPVFMRETSRLRYCDTFLGKTIAE-IPLFLVVFLEFETAIYAIPLIGRPGVDHFTALA 539
 QY 512 TLMWVAYSASSMALAIAGQSVSVATLMTTCFVMMIFSGLLVNLTTIASWLSWQYF 571

DB 540 LVTLVANVSTSGYLISCACSTSMALSVGPVPVPIIFLLFGGFFLNSGSPVAYFKWLSYL 599
 QY 572 SIPRYGPTALQHNELGQNFPG-LNATGNPCNVATCTGEEYLKQGLDLSPLWLNKH 630
 DB 600 SWFRYANEGLLINQW--ADVPRGEBITLSN-----TTCPSGSEVILETLNFSASGLPDDF 653
 QY 631 VALACMIVIFLITAYLKL 648
 DB 654 IGLALLIVGFRISAYIAL 671

 RESULT 13
 P91892
 ID P91892 PRELIMINARY; PRT; 692 AA.
 AC P91892;
 DT 01-MAY-1997 (TREMELrel. 03, Created)
 DT 01-MAY-1997 (TREMELrel. 03, Last sequence update)
 DT 01-MAY-2003 (TREMELrel. 23, Last annotation update)
 DE EYE pigment TRANSPORTER (EYE pigment transporter).
 GN WHITE.
 OS Aedes aegypti (Yellowfever mosquito).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Aedes.
 OX NCBI_TaxID=7159;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=ROCKEFELLER;
 RX MEDLINE=97418473; PubMed=9272447;
 RA Coates C.J., Schaub T.L., Besansky N.J., Collins F.H., James A.A.;
 RT "The white gene from the yellow fever mosquito, Aedes aegypti."
 RL Insect Mol. Biol. 6:291-299(1997).
 RN [2]
 RP SEQUENCE OF 1-20 FROM N.A.
 RC STRAIN=ROCKEFELLER;
 RA Coates C., Schaub T.L., James A.A.;
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
 DR EMBL; U88851; AAC04894.1; -.
 DR EMBL; AF051096; AAC05165.1; -.
 DR InterPro; IPR003593; AAA_ATPase.
 DR InterPro; IPR003439; ABC_transporter.
 DR InterPro; IPR005284; Pigment_permease.
 DR Pfam; PF00005; ABC_tran; 1.
 DR ProDom; PD000006; ABC transporter; 1.
 DR SMART; SM00382; AAA; 1.
 DR TIGRFAMS; TIGR00955; 3a01204; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
 KW ATP-binding; Transport.
 SQ SEQUENCE 692 AA; 77231 MW; DF2F00EEB4C17641 CRC64;

 Query Match 23.4%; Score 784; DB 5; Length 692;
 Best Local Similarity 32.0%; Pred. No. 3.9e-45;
 Matches 202; Conservative 123; Mismatches 258; Indels 48; Gaps 15;

 QY 42 ICYRV-----KLKSGFLPCPKVEKEILSNINGIMKPG-LNAILGPTGGKSSLDVLA 95
 DB 91 LCSRLCCCFRQKDFNP-----RKHLKNVTKMAKSGELLAVMGSSAGKTTLLNALSF 145
 QY 96 RKDPS---GLSGDVLINGAPRPA-NFKCNSGYVQDDVVMGTLTVRENLOFSAALRLATT 151
 DB 146 RSPGCVKIAPTSVRALNGIPVNAEQRLARCAVQDDDLFIPALTTRHLVPHAMLRMGKD 205
 QY 152 MTNHEKNERINRVIEELGLDKVADSKVGTQ-FIRGVSGGERKRTSIGMELITDPSILSD 210
 DB 206 VPKSVKMRNVNEVLQELSLAKCADIITIGAPORMKGLSGGERKRLAFASELTDPHLLCD 265
 QY 211 EPTTGLDSSSTANAVALLLXKMSKQGTITFISHOPRYSIFKLFDLSLTLASGRMLFHGPA 270
 DB 266 EPTGLDSSPMASVQLVQLKNGALGKTIITLHQPSSELYCLFDKILLVAGRVAFGLSP 325
 QY 271 QEALGYFESAGYHCEAYNNPADFFLDINGDSTAVALEEDFKAT-BIIFPSKQDKPLI 329
 DB 326 YQASEFFSQLGIPCPNPNYPADFYQVLT-----AIAPNKEACRDTIKKICDSFAVSPMA 380

```

QY 330 EKLAIIYVNSPYKETAELHOLSGEKKKIIIVFKBISYTTSFCHQLRWVSKRSFKNLL 389
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
381 REVMEV-ANSGKNVEEQYVLPMEGASR-----TGRSTWTFYVVLWRSWLTVL 430
QY 390 GNPOASIAQIIVTVVLGIVGAIYGLKNDSTGIONRAGVLFLTNQCFSSVSAV-ELF 448
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
431 KDPMLVKVRLLTANVATVLTGIIYFQRLDQDGVNNGALFLFLTNMTFQVFAVINVF 490
QY 449 VVEKKLFTHEYISGVYRVSSYFLGKLLSDLLPMRLPSIIFTCIVVFMGLKPKADAFV 508
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
491 SNAELPVFLREKRSRFRVDVTFLGKTIAR-VFLFLAVPFVFTSIYPMGLKSGATYILT 549
QY 509 MNFTLMVAYSASSMALAIAAQSVSVATLMTICFVFMIFSGLLVNLTTIASLWSL 568
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
550 ALLIVLVANVATSPCYLISCASSISMALSVGPPVLPFLIFGFFLNSASVPSYFVYL 609
QY 559 QYFSIPRGFTALQHNEFL-----GNQFCPLNATGNPCNATCTGEBYLVKQGLDLPW 624
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
610 SYFSWFRYANEALLINQWSTVOEGDIACTRANVT-----CPSSGQIILETFNFKVEDF 662
QY 625 GLWKHNVHALACMIVIFLTITAYLKLFLKKYS 655
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
663 GP---DIACLCLMIVIFLIGALFCLWLSRS 690

RESULT 14
Q9C6W5 PRELIMINARY; PRT; 648 AA.
AC Q9C6W5;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Hypothetical protein (ABC transporter, putative).
GN F2M3.2 OR AT1G31770/F2M3.2.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucots II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxId=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ecv. Columbia;
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA Dunn P., Etgu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.S., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Mafti R., Marziani A.,
RA Maltischer J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tanbarga G., Toriumi M.J., Town C.D.,
RA Utterback T., Van Aken S., Vayberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Xu G., Frazer C.M., Vencer J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
thaliana."
RL Nature 408:816-820(2000).
RP SEQUENCE FROM N.A.
RA Haas B.J., Volfovsky N., Town C.D., Trouthan M., Alexandrov N.,
RA Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;
RT "Full-length messenger RNA sequences greatly improve genome
annotation."
RL Genome Biol. 0:0-0(2002).
RN [3]
RP SEQUENCE FROM N.A.
RA Brover V., Trouthan M., Alexandrov N., Lu Y.-P., Flavell R.,

```

```

RA Feldmann K.;
RT "Full-length cDNA from Arabidopsis thaliana.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=ecv. Columbia;
RA Seki M., Iida K., Satou M., Sakurai T., Akiyama K., Ishida J.,
RA Nakajima M., Enju A., Kamiya A., Narusaka M., Carninci P., Kawai J.,
RA Hayashizaki Y., Shinozaki K.;
RT "Arabidopsis thaliana full-length cDNA.";
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC074360; AAG60152.1; -
DR EMBL; AY088793; AAM67104.1; -
DR EMBL; AK117530; BAC42192.1; -
DR InterPro; IPR003439; ABC transporter.
DR Pfam; PF00005; ABC tran_1.
DR ProDom; PD000006; ABC transporter; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW Hypothetical protein.
SQ SEQUENCE 648 AA; 72618 MW; D52A2D2434A5BB9D CRC64;

Query Match 23.3%; Score 781.5; DB 10; Length 648;
Best Local Similarity 33.5%; Pred. No. 5.3e-45;
Matches 219; Conservative 118; Mismatches 246; Indels 71; Gaps 23;

QY 20 FPATVSN-DLKAFTEGAVLSFHNICRYVKLK--SGFLPCRPVKEKILSNINGIMKPG-L 75
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
35 FPTITSQPLQMSMYPITLKEEVVYKVISETSOCMSWKSKEKTLINGITGMVPCGP 94
QY 76 NAILGPTGGKSSLLDVLAARKDPSGLSGDLINGAPRPNKNSGVYVQDDVVMGTILT 135
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
95 LAMLGPSSGKTLTLLSALGRLSKT-FSGKVMYNGQPSGCIKRTGTVAQDDVLYPHLT 153
QY 136 VRENLPESAALRLATTNTNHEKNERINRVIELGLDKVADSKVGTQFTIRGVSGGRKRTS 195
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
154 VWETLFFALLURLPSSLTRDEKAEHVDVIAELGNRTCSNMGIGLPRGSGGKKRVS 213
QY 196 IGMELITDPSILSDEPTTGLDSTANAVALLKRMKSQGRITIFSIHQPRYSIFKLFD 255
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
214 IGQEMLINPSSLLDLDEPTSLDSTTAHRIVTTIKRLASGGRTVVTTHIQSPSSRIYHMDK 273
QY 256 LTLASGLMHPGPAQEAALGYFESAGYHCEAYNNPADPFLDIING--DSTAVALNREDF 313
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
274 VVLLSEGPYIYGAASSAVEYFSSLGSTSLTVNPADILLDLANGIPDPDTOKETSEBQK 333
QY 314 KATRIEBSKODKPLEKL-AEI-YVNSSPYKETAELHOLSGEKKKIIIVFKBISYTT 371
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
334 TVKETL-VSAYEKNIISTKLKAEKCNASHSEYTKAAANKL-----KSEQWCT 380
QY 372 SFCHQL-----RWYKSKSFKNLLGNPQASIAQIIVTVVLGIVGAIYFGLKNDSTGIONR 426
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
381 TWWYQFTVLLQGVRRERFESF---NKLRIQVIVSAFLG---GLLWV--HTPKSHIQDR 432
QY 427 AGVLFFLTNQCPSV-SAVELFVVEKKLFTHEYISGVYRVSSYFLGKLLSDLLPMRLP 485
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
433 TALLFFSVFVGFVPLYNVATFPEQEKMLIKERSSGMYRLSSYFMAENVGD-LPLELAL 491
QY 486 SIIFTCIVVFMGLKPKADAFVMMFTLMVAYS---ASSMALAIAAQSVSVATLMT 542
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
492 PTAFVFIIYWMGGLKPDPTTF---ILSLVVLVYVLAQGLGAPGALLMMIKQATTLAS 548
QY 543 ICFVFMIFSGLLVNLTTIASLWSLQYFSIPRGFTALQHNEFLQGNFCPLNATGNP 602
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
549 VTLVFLIAGGYVQ-QIPPIVWLKYSYCYKLL-----GIQYTD-- 593
QY 603 CNYATCT-----GEEYLVK-QGIDLSPWGLKKNHVALACMIVIFLTITAYLKL 648
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
594 -DYVECKSGVWCRVGDPAIKSMGLN-----NLWIDVFMGVMLVGYRLMAYMAL 642

RESULT 15
Q9C6R7 PRELIMINARY; PRT; 646 AA.
ID Q9C6R7

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 1, 2003, 18:38:06 ; Search time 99.6384 Seconds
(without alignments)
1043.432 Million cell updates/sec

Title: US-09-856-927-2

Perfect score: 3350

Sequence: 1 MSSNVEFIPVSGQNTGPF.....MIVFLTIAYLKLLFLKKYS 655

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_13Jun03.*

1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	3350	100.0	655	AA199365	ATP-binding caset
2	3350	100.0	655	AA199365	Human BCRP/MXR/ABC
3	3350	100.0	655	AA199365	Homo sapiens ABC t
4	3350	100.0	655	AA199365	Human BCRP (huBCRP
5	3337	99.6	655	AA199365	Human transport pr
6	3337	99.6	655	AA199365	Human ABCG2. Homo
7	3337	99.6	655	AA199365	Human BCRP protein
8	3337	99.6	655	AA199365	Human ABCG2 mutant
9	3337	99.6	655	AA199365	Breast Cancer Resi

10	3330	99.4	665	23	AAO14783	Human BCRP-related
11	3329	99.4	655	23	ABO07273	Human BCRP (huBCRP
12	3322	99.2	665	23	AAO14782	Human BCRP-related
13	3047.5	91.0	604	23	AAW73627	Human secreted pro
14	3047.5	91.0	604	23	ABP61858	Human polypeptide
15	2754	82.2	657	23	ABO07272	Human BCRP (huBCRP
16	2323	69.3	456	22	AA093564	Human protein sequ
17	803	24.0	687	22	AB059384	Drosophila melanog
18	796	23.8	687	21	AA178981	Silkworm Bm white
19	781.5	23.3	625	21	AA180800	Arabidopsis thalia
20	781.5	23.3	632	21	AA180800	Arabidopsis thalia
21	781.5	23.3	648	21	AA180807	Arabidopsis thalia
22	743	22.2	649	21	AA180807	Arabidopsis thalia
23	743	22.2	653	21	AA180807	Arabidopsis thalia
24	724	21.6	832	22	AB059544	Drosophila melanog
25	717	21.4	616	21	AA050555	Arabidopsis thalia
26	712.5	21.3	665	23	AB057112	Mouse ischaemic co
27	708.5	21.1	674	23	ABP52126	Homo sapiens ABC t
28	706	21.1	638	23	AB098349	Human ABC transpor
29	705	21.0	646	23	AAE28968	Human ABCG4 transp
30	704	21.0	646	23	AB098348	Human ABC transpor
31	704	21.0	646	23	AAE28964	Human ABCG4 transp
32	702	21.0	646	24	AB098694	Amino acid sequenc
33	701	20.9	646	23	AAO14186	Human transporter
34	700.5	20.9	663	24	AB082647	Human DevG22 homol
35	694	20.7	646	24	AB098696	Amino acid sequenc
36	690.5	20.6	652	23	AAU96985	Mouse ABCG5 protei
37	689.5	20.6	652	23	AAE13289	Mouse sitosterolae
38	689.5	20.6	652	23	AAE13308	Mouse sitosterolae
39	689.5	20.6	652	23	AAE13702	Mouse ABCG5 protei
40	685	20.4	652	23	AAE13309	Mouse sitosterolae
41	678.5	20.3	652	23	AAU96986	Rat ABCG5 protei
42	674.5	20.1	651	23	AAU96984	Human ABCG5 protei
43	674.5	20.1	651	23	AAE13290	Human sitosterolae
44	674.5	20.1	651	24	AAE13704	Human ABCG5 protei
45	673.5	20.1	651	23	AAU96990	Human ABCG5 mutant

ALIGNMENTS

RESULT 1
AAU95365
ID AAU95365 standard; Protein; 655 AA.
AC AAU95365;
XX
XX
DT 25-SEP-2000 (first entry)
XX
DE ATP-binding cassette protein MXR1.
XX
XX
KW ATP-binding cassette protein; ABC protein; MXR1; human;
KW cytochrome resistance; colon carcinoma; therapy.
XX
XX
OS Homo sapiens.
PN WC000036101-A2.
XX
XX
PD 22-JUN-2000.
XX
XX
PF 24-NOV-1999; 99WO-US28107.
XX
XX
PR 30-NOV-1998; 98US-0110473.
XX
XX
XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
XX
XX Dean M, Allikmets R, Bates SE, Fojo AT;
XX
XX WPI; 2000-442385/38.
XX
XX N-PSDB; AAA27938.
XX
XX Adenosine triphosphate (ATP)-binding cassette proteins and their
XX nucleic acids, useful for identifying agents that may be used to treat
PT

PT colonic carcinomas resistant to cytotoxic therapeutic agents -
XX
PS Claim 1(ii); Page 46; 49pp; English.

XX The present sequence is that of an ATP-binding cassette protein
CC (ABC protein) designated MXR1 that confers mitoxantrone resistance
CC to S1-M1-80 human colon carcinoma cells. Nucleic acids encoding
CC the ABC protein may be used to produce the protein according to
CC standard recombinant DNA methodologies. The expressed proteins may
CC be used to confer resistance to cytotoxins such as mitoxantrone
CC and/or daunomycin to the host cells in which they are expressed.
CC In disease conditions such as carcinoma of the colon, breast and
CC gastrointestinal tract, the ABC protein may be expressed to provide
CC resistance to cytotoxic therapeutic agents. The nucleic acids and
CC proteins may be used to produce host cell models of resistant cells
CC which can be used to screen for candidate agents that inhibit the
CC expression and/or activity of the ABC protein. These agents may be
CC used to down regulate ABC protein expression in carcinoma cells and
CC sensitize them to cytotoxic therapeutic agents.

XX Sequence 655 AA;

Query Match 100.0%; Score 3350; DB 21; Length 655;
Best Local Similarity 100.0%; Pred. No. 4.6e-318;
Matches 655; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSSNVVEPIPVSQGNTNGFPATVNDLKAFTEGAVLSPHNICYRVKLSGFLPCRKPYE 60
DB 1 MSSNVVEPIPVSQGNTNGFPATVNDLKAFTEGAVLSPHNICYRVKLSGFLPCRKPYE 60
QY 61 KEILSNINGIMKPGNLAILGPTGGKSSLLDLAARKDPSGLSDVGLNGAPRANFKCN 120
DB 61 KEILSNINGIMKPGNLAILGPTGGKSSLLDLAARKDPSGLSDVGLNGAPRANFKCN 120
QY 121 SGYVQDDVVMGTLTVRENLOQSAALRLATTNNHEKNERINRVTEELGLDKVADSKVGT 180
DB 121 SGYVQDDVVMGTLTVRENLOQSAALRLATTNNHEKNERINRVTEELGLDKVADSKVGT 180
QY 181 QPIRGVSGGERKRTSIGMELITDPSILSLDEPTTGLDSTANAVLLLLKRSKQGRITIF 240
DB 181 QPIRGVSGGERKRTSIGMELITDPSILSLDEPTTGLDSTANAVLLLLKRSKQGRITIF 240
QY 241 SIHQPRYSIFKLFDSILTLASGLRMFHGPAQALGYFESAGYHCEAYNNPADFFLDIING 300
DB 241 SIHQPRYSIFKLFDSILTLASGLRMFHGPAQALGYFESAGYHCEAYNNPADFFLDIING 300
QY 301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAETIYVNSSFYKETAELHQLSGGEKKK 360
DB 301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAETIYVNSSFYKETAELHQLSGGEKKK 360
QY 361 ITVPKEISVTTSFCHOLRWVSKRSFKNLGNPQASIAQIIVTVLGLVIGALYFGLKND 420
DB 361 ITVPKEISVTTSFCHOLRWVSKRSFKNLGNPQASIAQIIVTVLGLVIGALYFGLKND 420
QY 421 TGIQNRAGVLPFLTTNQCFSSVSVELFVVEKKLFIEHYISGYRVSSYFGLKLSDLLP 480
DB 421 TGIQNRAGVLPFLTTNQCFSSVSVELFVVEKKLFIEHYISGYRVSSYFGLKLSDLLP 480
QY 481 NMLPSIITCTIYVPMGLKPADAFFVMFTLMVAVSASSMALAIAAGQSVSVATLL 540
DB 481 NMLPSIITCTIYVPMGLKPADAFFVMFTLMVAVSASSMALAIAAGQSVSVATLL 540
QY 541 MTICFVFMWIFSGLLVNLTTTASLWSLQYFSIPRYGFTALQHNFLGQNCPCGNATGN 600
DB 541 MTICFVFMWIFSGLLVNLTTTASLWSLQYFSIPRYGFTALQHNFLGQNCPCGNATGN 600
QY 601 NPCVATCTGEBYLVKQGDLSPLWGLKKNHVALACMIVIFTIAYKLLFLKKYS 655
DB 601 NPCVATCTGEBYLVKQGDLSPLWGLKKNHVALACMIVIFTIAYKLLFLKKYS 655

RESULT 2
AAU04348

ID AAU04348 standard; Protein; 655 AA.
XX
AC AAU04348;
XX
DT 23-OCT-2001 (first entry)
XX
DE Human BCRP/MXR/ABCP protein.
XX
KW ATP-binding cassette transporter protein-inhibiting peptide;
KW ABC; transmembrane domain; cancer; tumour; HIV; AIDS; BCRP/MXR/ABCP;
KW acquired immunodeficiency syndrome; human immunodeficiency syndrome.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Domain 398..420
FT /label= Transmembrane_domain_1
FT Domain 427..450
FT /label= Transmembrane_domain_2
FT Domain 478..501
FT /label= Transmembrane_domain_3
FT Domain 504..527
FT /label= Transmembrane_domain_4
FT Domain 539..558
FT /label= Transmembrane_domain_5
FT Domain 629..650
FT /label= Transmembrane_domain_6
XX
FN WO200136477-A2.
XX
PD 25-MAY-2001.
XX
PF 17-NOV-2000; 2000WO-US31817.
XX
PR 18-NOV-1999; 99US-0166382.
PR 22-NOV-1999; 99US-0166767.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Tarasova NI, Michejda CJ, Gottesman MM, Hrycyna C;
XX WPI; 2001-381224/40.
DR
XX New ATP-binding cassette transporter-inhibiting peptides derived from
XX transmembrane domain of the transporter, useful for inhibiting
XX biological activity of the transporter and sensitizing cancer cells to
XX chemotherapeutics -
XX
PS Disclosure; Fig 2; 89pp; English.
XX
CC The sequence represents human BCRP/MXR/ABCP protein, an ATP-binding
CC cassette (ABC) transporter protein. Peptides derived from its
CC transmembrane domains are used as ATP-binding cassette
CC transporter protein-inhibiting peptides. The peptides have a first end
CC and a second end, where at the first end there is a group that is
CC negatively charged under physiological conditions and at the second end a
CC group that is neutrally charged under physiological conditions and is at
CC least 70% identical to a transmembrane domain of an ABC transporter. The
CC peptides are used for inhibiting the biological activity, in particular
CC ion flux or translocation, cytotoxin efflux or translocation,
CC phosphorylation, protein synthesis or degradation, cellular morphology,
CC secretion, production of particular components such as soluble
CC inositol phosphates, tumour growth, chemotaxis, mitogenic response, cell
CC growth activation or secretion of a target ATP-binding cassette (ABC)
CC transporter. The peptides are useful to sensitize cancer
CC cells to standard chemotherapeutics, which chemotherapeutics are then
CC administered to kill the cancer cells. Transmembrane analogues of ABC
CC transporter protein are administered to patients infected with HIV-1,
CC who are taking a protease inhibitor. The ability of HIV-1 infected
CC cells to efflux the protease inhibitor is impaired, resulting in higher
CC concentration of the protease inhibitor in the affected cell. This
CC reduces the amount of protease inhibitor administered to the patient to
CC achieve a virus-inhibiting dose in the infected cells.


```
XX SQ Sequence 655 AA;
Query Match 100.0%; Score 3350; DB 22; Length 655;
Best Local Similarity 100.0%; Pred. No. 4.6e-318;
Matches 655; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSSNVVEFIPVSGQNTNGFPATVNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKPVE 60
DB 1 MSSNVVEFIPVSGQNTNGFPATVNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKPVE 60
QY 61 KEILSNINGIMKPGNAILGPTGGKSSLLDLAARKDPSGLSGDVLINGAPRANPKCN 120
DB 61 KEILSNINGIMKPGNAILGPTGGKSSLLDLAARKDPSGLSGDVLINGAPRANPKCN 120
QY 121 SGYVQDDVVMGTLTVRENLOQFSAAALRLATTMTNHEKNERINRVIEELGLDKVADSKVGT 180
DB 121 SGYVQDDVVMGTLTVRENLOQFSAAALRLATTMTNHEKNERINRVIEELGLDKVADSKVGT 180
QY 181 QFIRGVSGGERKRTSIGMELITDPSILSLDEPTTGLDSTANAVALLKRMKSQGRITIF 240
DB 181 QFIRGVSGGERKRTSIGMELITDPSILSLDEPTTGLDSTANAVALLKRMKSQGRITIF 240
QY 241 SIHQPRYSIFPKLFDLSLTLLASRLMFGHGAQALGFESAGYHCEAYNNPADFFLDIING 300
DB 241 SIHQPRYSIFPKLFDLSLTLLASRLMFGHGAQALGFESAGYHCEAYNNPADFFLDIING 300
QY 301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAIEYVNSSFYKETAELHQLSGGKKKK 360
DB 301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAIEYVNSSFYKETAELHQLSGGKKKK 360
QY 361 ITVFKESITTSFCHQLRWVSKRSFKNLLGNPQASIAQIIVTVVLGLVIGAIYFGLKND 420
DB 361 ITVFKESITTSFCHQLRWVSKRSFKNLLGNPQASIAQIIVTVVLGLVIGAIYFGLKND 420
QY 421 TGIQNRAGVLPFLTTNOCFSSVAVELFVVEKKLFIHEYISGYRVSSYFLGKLLSDLLP 480
DB 421 TGIQNRAGVLPFLTTNOCFSSVAVELFVVEKKLFIHEYISGYRVSSYFLGKLLSDLLP 480
QY 481 MRMLPSIIFTICIVYFMLGLPKADAFVMMFTLMWVAYSASSMALAIAGQSVSVATLL 540
DB 481 MRMLPSIIFTICIVYFMLGLPKADAFVMMFTLMWVAYSASSMALAIAGQSVSVATLL 540
```

RESULT 3

```
ABP52127
ID ABP52127 standard; Protein; 655 AA.
XX AC
XX ABP52127;
XX 10-OCT-2002 (first entry)
XX DE Homo sapiens ABC transporter ABCG2 protein SEQ ID NO:79.
```

```
XX KW ATP-binding cassette transporter; ABC transporter; modulation; D loop;
XX KW cancer; bacterial infection; fungal infection; protozoal infection;
XX KW antibacterial; fungicide; protozoaside.
```

```
XX OS Homo sapiens.
```

```
XX PN EP1217066-A1.
```

```
XX PD 26-JUN-2002.
```

```
XX PP 21-DEC-2000; 2000EP-0870316.
```

```
XX PR 21-DEC-2000; 2000EP-0870316.
XX PA (UYGE-) UNIV GENT.
XX WPI; 2002-550404/59.
XX
PT Modulating activity of ATP-binding cassette (ABC) transporters by
PT influencing dimerization of nucleotide binding domains through use of D
PT loop sequence of an ABC transporter, or its antisense peptide or
PT peptide mimetic -
XX
XX Disclosure; Fig 3; 290pp; English.
XX
CC The present invention describes a method (M1) for modulating the activity
CC of ATP-binding cassette (ABC) transporters by influencing the
CC dimerization of the nucleotide binding domains comprising: (a) a
CC polypeptide (polyP) consisting of 5-50 amino acids comprising the D loop
CC sequence of an ABC transporter (ABP52049 to ABP52091); (b) a polyP
CC consisting of the D loop sequence of an ABC transporter; (c) a peptide
CC mimetic or antisense peptide of (a) or (b). ABC transporters have
CC antibacterial, fungicide and protozoaside activities. (M1) is useful for
CC selectively modulating the activity of ABC transporters belonging to the
CC group of multidrug transporter/P-glycoproteins. Bacterial, fungal or
CC protozoal ABC transporters are involved in the infection of a mammal or
CC in the induction of resistance to antibiotics or drugs in a mammal. (M1)
CC is useful for preventing, treating or alleviating diseases associated
CC with functionality of an ABC transporter. ABP52092 to ABP52140 represent
CC ABC transporter proteins given in the exemplification of the present
CC invention.
```

Sequence 655 AA;

Query Match 100.0%; Score 3350; DB 23; Length 655;
Best Local Similarity 100.0%; Pred. No. 4.6e-318;
Matches 655; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MSSNVVEFIPVSGQNTNGFPATVNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKPVE 60
DB 1 MSSNVVEFIPVSGQNTNGFPATVNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKPVE 60
QY 61 KEILSNINGIMKPGNAILGPTGGKSSLLDLAARKDPSGLSGDVLINGAPRANPKCN 120
DB 61 KEILSNINGIMKPGNAILGPTGGKSSLLDLAARKDPSGLSGDVLINGAPRANPKCN 120
QY 121 SGYVQDDVVMGTLTVRENLOQFSAAALRLATTMTNHEKNERINRVIEELGLDKVADSKVGT 180
DB 121 SGYVQDDVVMGTLTVRENLOQFSAAALRLATTMTNHEKNERINRVIEELGLDKVADSKVGT 180
QY 181 QFIRGVSGGERKRTSIGMELITDPSILSLDEPTTGLDSTANAVALLKRMKSQGRITIF 240
DB 181 QFIRGVSGGERKRTSIGMELITDPSILSLDEPTTGLDSTANAVALLKRMKSQGRITIF 240
QY 241 SIHQPRYSIFPKLFDLSLTLLASRLMFGHGAQALGFESAGYHCEAYNNPADFFLDIING 300
DB 241 SIHQPRYSIFPKLFDLSLTLLASRLMFGHGAQALGFESAGYHCEAYNNPADFFLDIING 300
QY 301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAIEYVNSSFYKETAELHQLSGGKKKK 360
DB 301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAIEYVNSSFYKETAELHQLSGGKKKK 360
QY 361 ITVFKESITTSFCHQLRWVSKRSFKNLLGNPQASIAQIIVTVVLGLVIGAIYFGLKND 420
DB 361 ITVFKESITTSFCHQLRWVSKRSFKNLLGNPQASIAQIIVTVVLGLVIGAIYFGLKND 420
QY 421 TGIQNRAGVLPFLTTNOCFSSVAVELFVVEKKLFIHEYISGYRVSSYFLGKLLSDLLP 480
DB 421 TGIQNRAGVLPFLTTNOCFSSVAVELFVVEKKLFIHEYISGYRVSSYFLGKLLSDLLP 480
QY 481 MRMLPSIIFTICIVYFMLGLPKADAFVMMFTLMWVAYSASSMALAIAGQSVSVATLL 540
DB 481 MRMLPSIIFTICIVYFMLGLPKADAFVMMFTLMWVAYSASSMALAIAGQSVSVATLL 540
```

QY 541 MTICFVMMIFSGLLVNLTTIASWLSWLYFSIPRYGFTALQHNBEFLGQNFPCGLNATGN 600
 |||||
 Db 541 MTICFVMMIFSGLLVNLTTIASWLSWLYFSIPRYGFTALQHNBEFLGQNFPCGLNATGN 600
 |||||
 QY 601 NPCNVATCTGEEYLVKQGDLSFPGWLGKHNHVALACMIVIFLTIAVLKLLFLKKYS 655
 |||||
 Db 601 NPCNVATCTGEEYLVKQGDLSFPGWLGKHNHVALACMIVIFLTIAVLKLLFLKKYS 655
 |||||

RESULT 4

ABB07270
 ID ABB07270 standard; Protein; 655 AA.

XX
 AC ABB07270;

DT 26-MAR-2002 (first entry)

DE Human BCRP (huBCRP) amino acid sequence.

XX Stem cell; ATP transport protein; ATP-binding cassette; antiparkinsonian;
 KW hepatotropic; neurodegenerative; cytostatic; antianemic; muscular; BCRP;
 KW cardiant; gene therapy.

XX Homo sapiens.

XX WO200192877-A2.

XX 06-DEC-2001.

XX 30-MAY-2001; 2001WO-US17459.

XX 31-MAY-2000; 2000US-0584586.

XX 29-MAY-2001; 2001US-0866866.

XX (SJUD-) ST-JUDE CHILDREN'S RES HOSPITAL.

XX Sorrentino B, Schuetz J;

XX WPI; 2002-114368/15.

XX N-PSDB; ABA94369.

XX Identifying a stem cell, for treating e.g., muscular dystrophy,
 PT myocardial infarction, Parkinson's disease, or neurodegenerative
 PT disorders, comprises detecting the expression of an ATP transport
 PT protein (BCRP) by a cell

XX Examples; Page 78-80; 87pp; English.

XX The invention provides a method of identifying and/or isolating a stem
 CC cell that involves detecting the expression of an ATP transport protein
 CC containing a conserved ATP-binding cassette (BCRP) by a cell in a sample
 CC comprising stem cells. The isolated stem cells may be used in the
 CC treatment of diseases such as muscular dystrophy, degenerative liver
 CC disorder, myocardial infarction, Parkinson's disease, degenerative
 CC disorders of the brain, and for tissue regeneration or replacement.
 CC Haematopoietic cells can be used in bone marrow transplants (e.g., for
 CC treatment of leukemia) and for ex vivo gene therapy for treating blood
 CC diseases such as sickle cell anemia and thalassemia. The stem cells can
 CC also be used as cell targets in gene therapy protocols. The present
 CC sequence represents the amino acid sequence of human BCRP.

XX Sequence 655 AA;

Query Match 100.0%; Score 3350; DB 23; Length 655;

Best Local Similarity 100.0%; Pred. No. 4.6e-318;

Matches 655; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSSNVVEFIPVSQGNNGFPATVNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKPE 60

Db 1 MSSNVVEFIPVSQGNNGFPATVNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKPE 60

QY 61 KEILSNINGIMKPGNAILGPTGGKSLLDVLAARKDPSGLSGDVLINGAPRANFKCN 120
 |||||

Db 61 KEILSNINGIMKPGNAILGPTGGKSLLDVLAARKDPSGLSGDVLINGAPRANFKCN 120
 |||||
 QY 121 SGYVQDDVVMGTLTVRENLOFSALRLATTTMTHKNERINRVTEELGLDKVADSKVGT 180
 |||||
 Db 121 SGYVQDDVVMGTLTVRENLOFSALRLATTTMTHKNERINRVTEELGLDKVADSKVGT 180
 |||||
 QY 181 QFIRVSGGERKRTSIGMELITDPSILSLDPTTGLDSSSTANAVLLLLKRMASKQGRITIF 240
 |||||
 Db 181 QFIRVSGGERKRTSIGMELITDPSILSLDPTTGLDSSSTANAVLLLLKRMASKQGRITIF 240
 |||||
 QY 241 SIHQPRYSIFKLFDSLTLASRLMFGPAQEAALGYFESAGYHCEAYNPNADFFLDIING 300
 |||||
 Db 241 SIHQPRYSIFKLFDSLTLASRLMFGPAQEAALGYFESAGYHCEAYNPNADFFLDIING 300
 |||||
 QY 301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAIRIYVNSSFYKETAELHQLSGGEKKK 360
 |||||
 Db 301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAIRIYVNSSFYKETAELHQLSGGEKKK 360
 |||||
 QY 361 ITVFKESYTTSPFCHQLRWVSKRSFKNLLGNPQASIAQIIVTVVLGLVIGAIYFGLKNDS 420
 |||||
 Db 361 ITVFKESYTTSPFCHQLRWVSKRSFKNLLGNPQASIAQIIVTVVLGLVIGAIYFGLKNDS 420
 |||||
 QY 421 TGIQNRAGVLFLLTNNOCFSSVAVELFVVEKKLFIHEYISGYRVSSYFGLKLLSLLP 480
 |||||
 Db 421 TGIQNRAGVLFLLTNNOCFSSVAVELFVVEKKLFIHEYISGYRVSSYFGLKLLSLLP 480
 |||||
 QY 481 MRMLPSIIFTCIVYFMLGLKPKADAFVMMFTLMMVAYSASSMALATAAGOSVSVATLL 540
 |||||
 Db 481 MRMLPSIIFTCIVYFMLGLKPKADAFVMMFTLMMVAYSASSMALATAAGOSVSVATLL 540
 |||||
 QY 541 MTICFVMMIFSGLLVNLTTIASWLSWLYFSIPRYGFTALQHNBEFLGQNFPCGLNATGN 600
 |||||
 Db 541 MTICFVMMIFSGLLVNLTTIASWLSWLYFSIPRYGFTALQHNBEFLGQNFPCGLNATGN 600
 |||||
 QY 601 NPCNVATCTGEEYLVKQGDLSFPGWLGKHNHVALACMIVIFLTIAVLKLLFLKKYS 655
 |||||
 Db 601 NPCNVATCTGEEYLVKQGDLSFPGWLGKHNHVALACMIVIFLTIAVLKLLFLKKYS 655
 |||||

RESULT 5

AAB60104

ID AAB60104 standard; Protein; 655 AA.

XX AC AAB60104;

XX 28-MAR-2001 (first entry)

XX Human transport protein TPPT-24.

XX Human; transport protein; TPPT; transport disorder; metabolic disorder;
 KW neurological disorder; cardiovascular disorder; reproductive disorder;
 KW immune disorder; cancer.

XX Homo sapiens.

XX WO200078953-A2.

XX 28-DEC-2000.

XX 16-JUN-2000; 2000WO-US16668.

XX 17-JUN-1999; 99US-0139923.

XX 10-AUG-1999; 99US-0148177.

XX 18-AUG-1999; 99US-0149357.

XX 28-OCT-1999; 99US-0162287.

XX (INCY-) INCYTE GENOMICS INC.

XX Lal P, Yang J, Yue H, Hillman JL, Tang YT, Bandman O, Burford N;

XX Baughn MR, Azinzai Y, Lu DAM, Au-Young J, Patterson C;

XX WPI; 2001-041424/05.

XX N-PSDB; AAF27724.

XX Isolated polypeptide with a human transport protein sequence is useful
PT for the diagnosis, prevention and treatment of disorders associated
PT with the immune, reproductive and cardiovascular systems -
XX
XX
PS Claim 2; Page 126-127; 165pp; English.
XX
CC The present invention provides the protein and coding sequences for 43
CC novel human transport proteins (designated tppts). These can be used in
CC the diagnosis and treatment of transport, metabolic, neurological,
CC reproductive, cardiovascular and immune disorders, and cell proliferative
CC disorders such as cancer.
XX
XX Sequence 655 AA;

Query Match 99.6%; Score 3337; DB 22; Length 655;
Best Local Similarity 99.5%; Pred. No. 5.9e-317;
Matches 652; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MSSNVVEFIPVSGQNTNGFPATVNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKPYE 60
Db 1 MSSNVVEFIPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKPYE 60
QY 61 KEILSNINGIMKPGNAILGPTGGKSSLLDLVLAARKDPSGLSGDVLINGAPRPANFKCN 120
Db 61 KEILSNINGIMKPGNAILGPTGGKSSLLDLVLAARKDPSGLSGDVLINGAPRPANFKCN 120
QY 121 SGYVQDDVVMGTLTVRENLFQSAALRLATTMTNHEKNERINRVIEELGDKVADSKVGT 180
Db 121 SGYVQDDVVMGTLTVRENLFQSAALRLATTMTNHEKNERINRVIEELGDKVADSKVGT 180
QY 181 QFIRGVGGERKRTSIGMELITDPSILFDEPTTGLDSTANAVLLKRMKSKQGRITIF 240
Db 181 QFIRGVGGERKRTSIGMELITDPSILFDEPTTGLDSTANAVLLKRMKSKQGRITIF 240
QY 241 SIHQPRYSIFKLFDSLTLLASGRMLFHPGPAQALGFESAGYHCEAYNPNADFFLDIING 300
Db 241 SIHQPRYSIFKLFDSLTLLASGRMLFHPGPAQALGFESAGYHCEAYNPNADFFLDIING 300
QY 301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAIEIVNSSFYKETKAEHLHQLSGGKKK 360
Db 301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAIEIVNSSFYKETKAEHLHQLSGGKKK 360
QY 361 ITVPKEISYTTSFCHQLRWVSKRSFKNLLGNPQASIAQIIVTVVLGLVIGAIYFGLKND 420
Db 361 ITVPKEISYTTSFCHQLRWVSKRSFKNLLGNPQASIAQIIVTVVLGLVIGAIYFGLKND 420
QY 421 TGIQNRAGVLPFLTTNQCFSYSAVELFVVEKKLFIHEYISGYRVSSYFGLKLLSDLLP 480
Db 421 TGIQNRAGVLPFLTTNQCFSYSAVELFVVEKKLFIHEYISGYRVSSYFGLKLLSDLLP 480
QY 481 MRMLPSIIFTCTIVYFMLGLKPADAFFVNMFTLMWVAYSASSMALAIAAGQSVSVATLL 540
Db 481 MRMLPSIIFTCTIVYFMLGLKPADAFFVNMFTLMWVAYSASSMALAIAAGQSVSVATLL 540
QY 541 MTICFVMMIPSGLLVNLTTIASLWSLQYFSIPRYGFTALQHNFLQNFQCPGLNATGN 600
Db 541 MTICFVMMIPSGLLVNLTTIASLWSLQYFSIPRYGFTALQHNFLQNFQCPGLNATGN 600
QY 601 NPCNATCTGEEYLVKQIDLSFWGLWKNHVALACMIVIFLTIAVLLKLLFLKYS 655
Db 601 NPCNATCTGEEYLVKQIDLSFWGLWKNHVALACMIVIFLTIAVLLKLLFLKYS 655

RESULT 6

AAU80028

ID AAU80028 standard; Protein; 655 AA.

XX

XX

XX

DT 15-JUL-2002 (first entry)

XX

DE Human ABCG2.

XX Human; ABCG2; transporter protein; anticancer drug tolerance;
KW indocarbazole.
XX
OS Homo sapiens.
XX
PN WO200228894-A1.
XX
PD 11-APR-2002.
XX
PF 18-SEP-2001; 2001WO-JP08112.
XX
PR 03-OCT-2000; 2000JP-0303441.
XX
XX (BANY) BANYU PHARM CO LTD.
XX
XX Komatani H, Hara Y, Kotani H, Nakagawa R;
PI N-PSDB; ABK49901.
XX
DR WPI; 2002-352228/38.
XX
PT ABCG2 gene encoding transporter protein capable of selectively
PT transporting indocarbazole compounds, useful in screening inhibitors
PT and anticancer agents for administration in chemotherapy -
XX
PS Claim 1; Page 71-76; 98pp; Japanese.

CC The invention relates to an ABCG2 gene encoding a transporter protein
CC capable of imparting tolerance to an anticancer agent in mammals
CC comprising a fully defined sequence as given in the specification or an
CC amino acid sequence based on the sequence but with some amino acids
CC substituted, deleted or added. The gene and encoded protein are useful
CC in screening inhibitors and anticancer agents for administration in
CC chemotherapy with enhancement in sensitivity of cancer cell tolerance.
CC The gene relating to drug tolerance can be modified e.g. with the
CC transporter inhibitors, screened compounds, antibodies and antisense
CC nucleotides. The transporter is capable of selectively transporting
CC indocarbazole compounds extracellularly. The present sequence
CC represents the amino acid sequence of human ABCG2 protein.

SQ Sequence 655 AA;

Query Match 99.6%; Score 3337; DB 23; Length 655;
Best Local Similarity 99.5%; Pred. No. 5.9e-317;
Matches 652; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSSNVVEFIPVSGQNTNGFPATVNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKPYE 60
Db 1 MSSNVVEFIPVSGQNTNGFPATASNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKPYE 60
QY 61 KEILSNINGIMKPGNAILGPTGGKSSLLDLVLAARKDPSGLSGDVLINGAPRPANFKCN 120
Db 61 KEILSNINGIMKPGNAILGPTGGKSSLLDLVLAARKDPSGLSGDVLINGAPRPANFKCN 120
QY 121 SGYVQDDVVMGTLTVRENLFQSAALRLATTMTNHEKNERINRVIEELGDKVADSKVGT 180
Db 121 SGYVQDDVVMGTLTVRENLFQSAALRLATTMTNHEKNERINRVIEELGDKVADSKVGT 180
QY 181 QFIRGVGGERKRTSIGMELITDPSILFDEPTTGLDSTANAVLLKRMKSKQGRITIF 240
Db 181 QFIRGVGGERKRTSIGMELITDPSILFDEPTTGLDSTANAVLLKRMKSKQGRITIF 240
QY 241 SIHQPRYSIFKLFDSLTLLASGRMLFHPGPAQALGFESAGYHCEAYNPNADFFLDIING 300
Db 241 SIHQPRYSIFKLFDSLTLLASGRMLFHPGPAQALGFESAGYHCEAYNPNADFFLDIING 300
QY 301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAIEIVNSSFYKETKAEHLHQLSGGKKK 360
Db 301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAIEIVNSSFYKETKAEHLHQLSGGKKK 360
QY 361 ITVPKEISYTTSFCHQLRWVSKRSFKNLLGNPQASIAQIIVTVVLGLVIGAIYFGLKND 420
Db 361 ITVPKEISYTTSFCHQLRWVSKRSFKNLLGNPQASIAQIIVTVVLGLVIGAIYFGLKND 420

```
QY 421 TGIQNRAGVLPFLTTNOCFSSVSAVELFWVEKKLFIHEYISGYRVSSYFLGKLLSDLLP 480
DB 421 TGIQNRAGVLPFLTTNOCFSSVSAVELFWVEKKLFIHEYISGYRVSSYFLGKLLSDLLP 480
QY 481 MRMLPSIIITCIVYFMGLGKPKADAFFVMMFTLMVAYSASSMALAIAAGQSVSVATLL 540
DB 481 MRMLPSIIITCIVYFMGLGKPKADAFFVMMFTLMVAYSASSMALAIAAGQSVSVATLL 540
QY 541 MTICFVFMIFSGLLVNLTTIASWLSWLOQYFSIPRYGFTALQHNFLGQFCPLGNATGN 600
DB 541 MTICFVFMIFSGLLVNLTTIASWLSWLOQYFSIPRYGFTALQHNFLGQFCPLGNATGN 600
QY 601 NPCNATCTGEEYLVKQIDLSPWGLWKNHVALACMIVIFLTIAYLKLLFLKKYS 655
DB 601 NPCNATCTGEEYLVKQIDLSPWGLWKNHVALACMIVIFLTIAYLKLLFLKKYS 655

RESULT 7
AAO14781
ID AAO14781 standard; Protein; 655 AA.
XX
AC AAO14781;
XX
DT 28-JUN-2002 (first entry)
XX
DE Human BCRP protein.
XX
KW Human; BCRP protein; membrane penetrating region; cancer.
XX
OS Homo sapiens.
XX
PN JP2002065277-A.
XX
PD 05-MAR-2002.
XX
PF 31-AUG-2000; 2000JP-0263742.
XX
PR 31-AUG-2000; 2000JP-0263742.
XX
PA (GANK-) ZH GAN KENYUKAI.
XX
DR WPI; 2002-324198/36.
XX
DR N-PSDB; AAL42412.
XX
PT Mutant BCRP protein useful for treatment of cancer -
XX
PS Claim 13; Page 7-8; 15pp; Japanese.
XX
CC The invention comprises a mutant human BCRP protein, having a deletion,
CC replacement or addition of at least one amino acid in the fifth membrane
CC penetrating region of the wild-type BCRP protein. The mutant BCRP protein
CC can be used for the treatment of cancer. The present amino acid sequence
CC represents a human BCRP protein.
XX
SQ Sequence 655 AA;

Query Match 99.6%; Score 3337; DB 23; Length 655;
Best Local Similarity 99.5%; Pred. No. 5.9e-317;
Matches 652; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSSNVEFIPVSCQNTNGFPATVNDLKAFTEGAVLSFHNICVYRVKLSGFLPCRKPE 60
DB 1 MSSNVEFIPVSCQNTNGFPATVNDLKAFTEGAVLSFHNICVYRVKLSGFLPCRKPE 60
QY 61 KEILSNINGIMKPLGNAILGPTGGKSSILLDLVLAARKDPSGLSGDLVINGAPRANFKCN 120
DB 61 KEILSNINGIMKPLGNAILGPTGGKSSILLDLVLAARKDPSGLSGDLVINGAPRANFKCN 120
QY 121 SGYVVQDDVMGTLTVRENLOFSAALRLATWTNHEKNERINRVIEELGLDKVADSKVGT 180
DB 121 SGYVVQDDVMGTLTVRENLOFSAALRLATWTNHEKNERINRVIEELGLDKVADSKVGT 180
```

```
QY 181 QFIRGVSGGERKRTSIGMELITDPSILSLDEPTTGLDSSSTANAVALLLKRMKSKQGRITIF 240
DB 181 QFIRGVSGGERKRTSIGMELITDPSILSLDEPTTGLDSSSTANAVALLLKRMKSKQGRITIF 240
QY 241 SIHQPRYSIFKLPDSLTLASGRLMFHGPQAQALGYFESAGYHCEAYNNPADFFLDIING 300
DB 241 SIHQPRYSIFKLPDSLTLASGRLMFHGPQAQALGYFESAGYHCEAYNNPADFFLDIING 300
QY 301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAIIYNNSSFYKTKAELHQLSGEKKKK 360
DB 301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAIIYNNSSFYKTKAELHQLSGEKKKK 360
QY 361 ITVFKELSYTTSFCHQLRWVSKRSFKNLLGNPQASIAQIIVTVVLGLVIGAIYFGLKNDS 420
DB 361 ITVFKELSYTTSFCHQLRWVSKRSFKNLLGNPQASIAQIIVTVVLGLVIGAIYFGLKNDS 420
QY 421 TGIQNRAGVLPFLTTNOCFSSVSAVELFWVEKKLFIHEYISGYRVSSYFLGKLLSDLLP 480
DB 421 TGIQNRAGVLPFLTTNOCFSSVSAVELFWVEKKLFIHEYISGYRVSSYFLGKLLSDLLP 480
QY 481 MRMLPSIIITCIVYFMGLGKPKADAFFVMMFTLMVAYSASSMALAIAAGQSVSVATLL 540
DB 481 MRMLPSIIITCIVYFMGLGKPKADAFFVMMFTLMVAYSASSMALAIAAGQSVSVATLL 540
QY 541 MTICFVFMIFSGLLVNLTTIASWLSWLOQYFSIPRYGFTALQHNFLGQFCPLGNATGN 600
DB 541 MTICFVFMIFSGLLVNLTTIASWLSWLOQYFSIPRYGFTALQHNFLGQFCPLGNATGN 600
QY 601 NPCNATCTGEEYLVKQIDLSPWGLWKNHVALACMIVIFLTIAYLKLLFLKKYS 655
DB 601 NPCNATCTGEEYLVKQIDLSPWGLWKNHVALACMIVIFLTIAYLKLLFLKKYS 655

RESULT 8
AAU80029
ID AAU80029 standard; Protein; 655 AA.
XX
AC AAU80029;
XX
DT 15-JUL-2002 (first entry)
XX
DE Human ABCG2 mutant 482T.
XX
KW Human; ABCG2; transporter protein; anticancer drug tolerance;
KW indocarbazole; mutant; muten.
XX
OS Homo sapiens.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 482
FT /note= "Wild type Arg substituted by Thr"
XX
PN WO200228894-A1.
XX
PD 11-APR-2002.
XX
PF 18-SEP-2001; 2001WO-JP08112.
XX
PR 03-OCT-2000; 2000JP-0303441.
XX
PA (BANY ) BANYU PHARM CO LTD.
XX
PI Komatani H, Hara Y, Kotani H, Nakagawa R;
XX
DR WPI; 2002-352228/38.
XX
DR N-PSDB; ABK49911.
XX
PT ABCG2 gene encoding transporter protein capable of selectively
PT transporting indocarbazole compounds, useful in screening inhibitors
XX and anticancer agents for administration in chemotherapy -
PS Disclosure; Page 87-90; 98pp; Japanese.
```

XX The invention relates to an ABCG2 gene encoding a transporter protein
 CC capable of imparting tolerance to an anticancer agent in mammals
 CC comprising a fully defined sequence as given in the specification or an
 CC amino acid sequence based on the sequence but with some amino acids
 CC substituted, deleted or added. The gene and encoded protein are useful
 CC in screening inhibitors and anticancer agents for administration in
 CC chemotherapy with enhancement in sensitivity of cancer cell tolerance.
 CC The gene relating to drug tolerance can be modified e.g. with the
 CC transporter inhibitors, screened compounds, antibodies and antisense
 CC nucleotides. The transporter is capable of selectively transporting
 CC indocarbazole compounds extracellularly. The present sequence
 CC represents the amino acid sequence of human ABCG2 mutant 482T.
 XX
 SQ Sequence 655 AA;
 Query Match 99.4%; Score 3331; DB 23; Length 655;
 Best Local Similarity 99.4%; Pred. No. 2.3e-316;
 Matches 651; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 1 MSSSNVEFIPVSGQNTNGFPATVSNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKPVE 60
 DB 1 MSSSNVEFIPVSGQNTNGFPATASNDLKAFTGAVLSFHNICYRVKLSGFLPCRKPVE 60
 QY 61 KEILSNINGIMKPLNALIGPTGGKSSLLDLVLAARKDPSGLSGDVLINGAPRANFKCN 120
 DB 61 KEILSNINGIMKPLNALIGPTGGKSSLLDLVLAARKDPSGLSGDVLINGAPRANFKCN 120
 QY 121 SGYVQDDVVMGTLTVRENLOFSALRLATTNTNHEKNERINRVTEELGDKVADSKVGT 180
 DB 121 SGYVQDDVVMGTLTVRENLOFSALRLATTNTNHEKNERINRVTEELGDKVADSKVGT 180
 QY 181 QFIRGVSGGERKRTSIGMBLITDPSILSDPTGLDSTANAVALLLRMSKQGRITIF 240
 DB 181 QFIRGVSGGERKRTSIGMBLITDPSILFDEPTGLDSTANAVALLLRMSKQGRITIF 240
 QY 241 SIHQPRYSIFKLFDSLTLLASGRMLPHGPAQALGYFESAGYHCEAYNNPADFFLDIING 300
 DB 241 SIHQPRYSIFKLFDSLTLLASGRMLPHGPAQALGYFESAGYHCEAYNNPADFFLDIING 300
 QY 301 DSTAVALNREEDFKATEIIEPSKODKPLIEKLAIEIVNSSFYKETKAEHLQLSGGEKKK 360
 DB 301 DSTAVALNREEDFKATEIIEPSKODKPLIEKLAIEIVNSSFYKETKAEHLQLSGGEKKK 360
 QY 361 ITVFKEISYTTSFCHQLRWVSKRSFKNLLGNQOASIAQIIVTVLGLVIGALYFGLKND 420
 DB 361 ITVFKEISYTTSFCHQLRWVSKRSFKNLLGNQOASIAQIIVTVLGLVIGALYFGLKND 420
 QY 421 TGIONRAGVLFPLTTNQCFSSVSVELFVVEKKLPIHEIYISGYRVSSYFLGKLLSDLLP 480
 DB 421 TGIONRAGVLFPLTTNQCFSSVSVELFVVEKKLPIHEIYISGYRVSSYFLGKLLSDLLP 480
 QY 481 NMLPSIIITCIVYFMGLGPKPADAFFVMFTLMVAYSASSMALAIAAGQSVSVATLL 540
 DB 481 NMLPSIIITCIVYFMGLGPKPADAFFVMFTLMVAYSASSMALAIAAGQSVSVATLL 540
 QY 541 MTICFVFMIFSGLLVNLTTIASWSLQYFSIPRYGFTALOHNEFLGQFCPLNATGN 600
 DB 541 MTICFVFMIFSGLLVNLTTIASWSLQYFSIPRYGFTALOHNEFLGQFCPLNATGN 600
 QY 601 NPCNATCTGEEYLVKQGDIDLPWGLKNHVALACMIVIFLTAYLKLFLKKYS 655
 DB 601 NPCNATCTGEEYLVKQGDIDLPWGLKNHVALACMIVIFLTAYLKLFLKKYS 655

RESULT 9

AA15221

ID AA15221 standard; Protein; 663 AA.

XX AA15221;

XX 09-NOV-1999 (first entry)

XX

DE Breast Cancer Resistance Protein (BCRP).
 XX breast cancer; drug resistance; ATP-binding cassette; ABC;
 KW xenobiotic transporter; chemotherapy; mitoxantrone; doxorubicin;
 KW breast cancer resistance protein; BCRP.
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Domain 87..95
 FT Domain /note= "Walker A motif"
 FT Domain 221..236
 FT Modified-site /note= "Phosphopantetheine site"
 FT Modified-site 345..347
 FT Region /note= "Glycosylation site on N"
 FT Region 405..422
 FT /label= TM1
 FT Modified-site /note= "Transmembrane region"
 FT Modified-site 425..427
 FT Region /note= "Glycosylation site on N"
 FT Region 546..563
 FT Modified-site /label= TM2
 FT Modified-site 564..566
 FT Modified-site /note= "Glycosylation site on N"
 FT Modified-site 604..606
 FT Region /note= "Glycosylation site on N"
 FT Region 638..655
 FT /label= TM3
 XX
 PN W09940110-A1
 XX 12-AUG-1999.
 PD
 XX 05-FEB-1999; 99WO-US02577.
 PF
 XX 05-FEB-1998; 98US-0073763.
 PR (UYMA-) UNIV MARYLAND BALTIMORE.
 PA
 XX Abruzzo L, Doyle LA, Ross DD;
 XX WPI; 1999-494273/41.
 DR N-PSDB; AA206360.
 XX
 PT New breast cancer resistance protein useful for production of
 PT antibodies to inhibit resistance activity for enhancing chemotherapy
 PT treatment
 XX
 PS Claim 4; Fig 2a; 80pp; English.
 XX
 CC The Breast Cancer Resistance Protein (BCRP) is an ATP-binding cassette
 CC (ABC) transporter protein. It has a molecular mass of approximately
 CC 72.3 kilodaltons (kD) exclusive of any glycosylation. Expression of BCRP
 CC in drug sensitive human cancer cells confers resistance to mitoxantrone,
 CC doxorubicin, and daunorubicin, and reduces daunorubicin accumulation in
 CC the cloned transfected cells.
 CC The protein is useful for producing antibodies and antisense
 CC probes, which can be used to inhibit the activity of BCRP, therefore
 CC enhancing a cancer patient's chemotherapy treatment.
 CC The antibodies and probes overcome the problems of breast cancer
 CC resistance proteins to make chemotherapy treatment more
 CC effective.
 XX
 SQ Sequence 663 AA;
 Query Match 99.4%; Score 3331; DB 20; Length 663;
 Best Local Similarity 99.4%; Pred. No. 2.3e-316;
 Matches 651; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 1 MSSSNVEFIPVSGQNTNGFPATVSNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKPVE 60
 DB 9 MSSSNVEFIPVSGQNTNGFPATASNDLKAFTGAVLSFHNICYRVKLSGFLPCRKPVE 68

QY 61 KEILSNINGIMKPLNAILGPTGGKSSLLDLVLAARKDPSGLSGDVLINGAPRANFKCN 120
DB 69 KEILSNINGIMKPLNAILGPTGGKSSLLDLVLAARKDPSGLSGDVLINGAPRANFKCN 128
QY 121 SGYVQDDVVMGTLTVRENLOFSAALRLATTMTNHEKNERINRVIELGLDKVADSKVGT 180
DB 129 SGYVQDDVVMGTLTVRENLOFSAALRLATTMTNHEKNERINRVIELGLDKVADSKVGT 188
QY 181 QFIRGVSGGERKRTSIGMELITDPSILSLDEPTTGLDSTANAVALLKRMKSKQGRITIF 240
DB 189 QFIRGVSGGERKRTSIGMELITDPSILSLDEPTTGLDSTANAVALLKRMKSKQGRITIF 248
QY 241 SIHQPRYSIFKLFDSJTLASGLRMFHPGPAQEAALGYFESAGYHCEAYNNPADFFLDIING 300
DB 249 SIHQPRYSIFKLFDSJTLASGLRMFHPGPAQEAALGYFESAGYHCEAYNNPADFFLDIING 308
QY 301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAIEIYVNSFYKETKAEHLQSGGKKKK 360
DB 309 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAIEIYVNSFYKETKAEHLQSGGKKKK 368
QY 361 ITVFEKISYTTSFCHOLRWVSKSPKLLGNPOASIAQIIVTVVLGLVIGAIYFGLKND 420
DB 369 ITVFEKISYTTSFCHOLRWVSKSPKLLGNPOASIAQIIVTVVLGLVIGAIYFGLKND 428
QY 421 TGIQNRAGVLFLLTNQCFSSSAVELFVVEKKLFIHEIYISGYRVSSYFGLKLLSDLLP 480
DB 429 TGIQNRAGVLFLLTNQCFSSSAVELFVVEKKLFIHEIYISGYRVSSYFGLKLLSDLLP 488
QY 481 MRMLPSIIIFTICIVFMLGLKPKADAFVMMFTLMVAYSSASSMALAIAAGQSVSVATLL 540
DB 489 MTMLPSIIIFTICIVFMLGLKPKADAFVMMFTLMVAYSSASSMALAIAAGQSVSVATLL 548
QY 541 MTICFVFMIFSGLLVNLTTIASWLSWLOFSPRYGFTALQHNELQHNELQHNELQHNEL 600
DB 549 MTICFVFMIFSGLLVNLTTIASWLSWLOFSPRYGFTALQHNELQHNELQHNELQHNEL 608
QY 601 NPCNYATCTGEEYLVKQIDLSPLWGLWKNHVALACMIVIFLTIAVILKLLFLKKYS 655
DB 609 NPCNYATCTGEEYLVKQIDLSPLWGLWKNHVALACMIVIFLTIAVILKLLFLKKYS 663

RESULT 10

AAO14783
ID AAO14783 standard; Protein; 665 AA.

AC AAO14783;

XX 28-JUN-2002 (first entry)

DT Human BCRP-related protein 2.

DE Human; BCRP-related protein; membrane penetrating region; cancer.

KW Homo sapiens.

OS JP2002065277-A.

XX 05-MAR-2002.

XX 31-AUG-2000; 2000JP-0263742.

XX 31-AUG-2000; 2000JP-0263742.

XX (GANK-) ZH GAN KENYUKAI.

XX WPI; 2002-324198/36.

DR N-PSDB; AAL42414.

XX Mutant BCRP protein useful for treatment of cancer -

XX Example 3; Page 12-13; 15pp; Japanese.

PS The invention comprises a mutant human BCRP protein, having a deletion,

CC replacement or addition of at least one amino acid in the fifth membrane
CC penetrating region of the wild-type BCRP protein. The mutant BCRP protein
CC can be used for the treatment of cancer. The present amino acid sequence
CC represents a human BCRP-related protein.

XX Sequence 665 AA;

Query Match 99.4%; Score 3330; DB 23; Length 665;

Best Local Similarity 99.4%; Pred. No. 2.9e-316; Mismatches 3; Indels 0; Gaps 0;
Matches 651; Conservative 1;

QY 1 MSSNVVEFIPVQSGNTNGFPATVNDLKAFTGAVLSFHNICYRVKLSGFLPCRKPYE 60
DB 11 MSSNVVEFIPVQSGNTNGFPATVNDLKAFTGAVLSFHNICYRVKLSGFLPCRKPYE 70
QY 61 KEILSNINGIMKPLNAILGPTGGKSSLLDLVLAARKDPSGLSGDVLINGAPRANFKCN 120
DB 71 KEILSNINGIMKPLNAILGPTGGKSSLLDLVLAARKDPSGLSGDVLINGAPRANFKCN 130
QY 121 SGYVQDDVVMGTLTVRENLOFSAALRLATTMTNHEKNERINRVIELGLDKVADSKVGT 180
DB 131 SGYVQDDVVMGTLTVRENLOFSAALRLATTMTNHEKNERINRVIELGLDKVADSKVGT 190
QY 181 QFIRGVSGGERKRTSIGMELITDPSILSLDEPTTGLDSTANAVALLKRMKSKQGRITIF 240
DB 191 QFIRGVSGGERKRTSIGMELITDPSILSLDEPTTGLDSTANAVALLKRMKSKQGRITIF 250
QY 241 SIHQPRYSIFKLFDSJTLASGLRMFHPGPAQEAALGYFESAGYHCEAYNNPADFFLDIING 300
DB 251 SIHQPRYSIFKLFDSJTLASGLRMFHPGPAQEAALGYFESAGYHCEAYNNPADFFLDIING 310
QY 301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAIEIYVNSFYKETKAEHLQSGGKKKK 360
DB 311 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAIEIYVNSFYKETKAEHLQSGGKKKK 370
QY 361 ITVFEKISYTTSFCHOLRWVSKSPKLLGNPOASIAQIIVTVVLGLVIGAIYFGLKND 420
DB 371 ITVFEKISYTTSFCHOLRWVSKSPKLLGNPOASIAQIIVTVVLGLVIGAIYFGLKND 430
QY 421 TGIQNRAGVLFLLTNQCFSSSAVELFVVEKKLFIHEIYISGYRVSSYFGLKLLSDLLP 480
DB 431 TGIQNRAGVLFLLTNQCFSSSAVELFVVEKKLFIHEIYISGYRVSSYFGLKLLSDLLP 490
QY 481 MRMLPSIIIFTICIVFMLGLKPKADAFVMMFTLMVAYSSASSMALAIAAGQSVSVATLL 540
DB 491 MRMLPSIIIFTICIVFMLGLKPKADAFVMMFTLMVAYSSASSMALAIAAGQSVSVATLL 550
QY 541 MTICFVFMIFSGLLVNLTTIASWLSWLOFSPRYGFTALQHNELQHNELQHNELQHNEL 600
DB 551 MTICFVFMIFSGLLVNLTTIASWLSWLOFSPRYGFTALQHNELQHNELQHNELQHNEL 610
QY 601 NPCNYATCTGEEYLVKQIDLSPLWGLWKNHVALACMIVIFLTIAVILKLLFLKKYS 655
DB 611 NPCNYATCTGEEYLVKQIDLSPLWGLWKNHVALACMIVIFLTIAVILKLLFLKKYS 665

RESULT 11

ABB07273
ID ABB07273 standard; Protein; 655 AA.

XX ABB07273;

XX 26-MAR-2002 (first entry)

XX Human BCRP (huBCRP) sequence.

DE Stem cell; ATP transport protein; ATP-binding cassette; antiparkinsonian;
KW hepatotropic; neurodegenerative; cytostatic; antianemic; muscular; BCRP;
KW cardiant; gene therapy.

XX Homo sapiens.

XX WO200192877-A2.

XX 06-DEC-2001.
XX 30-MAY-2001; 2001WO-US17459.
XX 31-MAY-2000; 2000US-0584586.
XX 29-MAY-2001; 2001US-0866866.
XX (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.
XX Sorrentino B, Schuetz J;
XX WPI; 2002-114368/15.
XX N-PSDB; ABA94383.
XX Identifying a stem cell, for treating e.g., muscular dystrophy,
XX myocardial infarction, Parkinson's disease, or neurodegenerative
XX disorders, comprises detecting the expression of an ATP transport
XX protein (BCRP) by a cell -
XX Example 1; Page 86-87; 87pp; English.
XX The invention provides a method of identifying and/or isolating a stem
XX cell that involves detecting the expression of an ATP transport protein
XX containing a conserved ATP-binding cassette (BCRP) by a cell in a sample
XX comprising stem cells. The isolated stem cells may be used in the
XX treatment of diseases such as muscular dystrophy, degenerative liver
XX disorder, myocardial infarction, Parkinson's disease, degenerative
XX disorders of the brain, and for tissue regeneration or replacement.
XX Haematopoietic cells can be used in bone marrow transplants (e.g., for
XX treatment of leukemia) and for ex vivo gene therapy for treating blood
XX diseases such as sickle cell anemia and thalassemia. The stem cells can
XX also be used as cell targets in gene therapy protocols. The present
XX sequence represents the human BCRP.
XX Sequence 655 AA;
Query Match 99.4%; Score 3329; DB 23; Length 655;
Best Local Similarity 99.4%; Pred. No. 3.6e-316;
Matches 651; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 1 MSSNVEVPIPVSOQNTNGFPATVNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKPVE 60
DB 1 MSSNVEVPIPVSOQNTNGFPATVNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKPVE 60
QY 61 KEILSNINGIMKPGNLAILGPTGGKSLDLVLAARKDPSGLSGDVLINGAPRPNFKCN 120
DB 61 KEILSNINGIMKPGNLAILGPTGGKSLDLVLAARKDPSGLSGDVLINGAPRPNFKCN 120
QY 121 SGYVQDDVVMGTLTVRENLOFSALRLATTMTNHEKNERINRVIEELGLDKVADSKVGT 180
DB 121 SGYVQDDVVMGTLTVRENLOFSALRLATTMTNHEKNERINRVIEELGLDKVADSKVGT 180
QY 181 QIRGVSGGERKRTSIGMELITDPSILSDPTGLDSTANAVALLKRMKSKQGRRTIIF 240
DB 181 QIRGVSGGERKRTSIGMELITDPSILSDPTGLDSTANAVALLKRMKSKQGRRTIIF 240
QY 241 SIHQPRYSIFKLFDSTLLASGRMLPHGPAQALGYFESAGYHCEAYNNPADFFLDIING 300
DB 241 SIHQPRYSIFKLFDSTLLASGRMLPHGPAQALGYFESAGYHCEAYNNPADFFLDIING 300
QY 301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAIEIVNSSFYKETKAEHLQSLGGKSKK 360
DB 301 DSTAVALNREEDFKATEIIEPSKQDKPLIEKLAIEIVNSSFYKETKAEHLQSLGGKSKK 360
QY 361 ITVFEISYTSFCHQLRWVSKRSPKLLGNPOASIAQIIVTVGLVIGAIYFGLKND 420
DB 361 ITVFEISYTSFCHQLRWVSKRSPKLLGNPOASIAQIIVTVGLVIGAIYFGLKND 420
QY 421 TGIONRAGVLFLLTNNQCFSSYSAVELFVVEKKLFIHEISGYRYVSSYFLGKLLSDLLP 480
DB 421 TGIONRAGVLFLLTNNQCFSSYSAVELFVVEKKLFIHEISGYRYVSSYFLGKLLSDLLP 480

QY 481 MRMLPSIIPTCIVYFMLGLKPKADAFFVMMFTLMMVAYSASSMALAIAGQSVSVATLL 540
DB 481 MRMLPSIIPTCIVYFMLGLKPKADAFFVMMFTLMMVAYSASSMALAIAGQSVSVATLL 540
QY 541 MTICFVFMWIFSGLLVNLTTIASWLSWLOYSIPRYGFTALQHNFLGONFCPGLNATGN 600
DB 541 MTICFVFMWIFSGLLVNLTTIASWLSWLOYSIPRYGFTALQHNFLGONFCPGLNATGN 600
QY 601 NPCNYATCTGBEYLKQIGIDLSPMGLWKNHVALACMIVIFLTIAYLKLLFLKKYS 655
DB 601 NPCNYATCTGBEYLKQIGIDLSPMGLWKNHVALACMIVIFLTIAYLKLLFLKKYS 655
RESULT 12
AAO14782
ID AAO14782 standard; Protein; 665 AA.
AC AAO14782;
XX 28-JUN-2002 (first entry)
XX Human BCRP-related protein 1.
XX Human; BCRP-related protein; membrane penetrating region; cancer.
XX Homo sapiens.
XX JP2002065277-A.
XX 05-MAR-2002.
XX 31-AUG-2000; 2000JP-0263742.
XX 31-AUG-2000; 2000JP-0263742.
XX (GANK-) ZH GAN KENYUKAI.
XX WPI; 2002-324198/36.
XX N-PSDB; AAL42413.
XX Mutant BCRP protein useful for treatment of cancer -
XX Example 2; Page 9-11; 15pp; Japanese.
XX The invention comprises a mutant human BCRP protein, having a deletion,
XX replacement or addition of at least one amino acid in the fifth membrane
XX penetrating region of the wild-type BCRP protein. The mutant BCRP protein
XX can be used for the treatment of cancer. The present amino acid sequence
XX represents a human BCRP-related protein.
XX Sequence 665 AA;
Query Match 99.2%; Score 3322; DB 23; Length 665;
Best Local Similarity 99.1%; Pred. No. 1.8e-315;
Matches 649; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
QY 1 MSSNVEVPIPVSOQNTNGFPATVNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKPVE 60
DB 11 MSSNVEVPIPVSOQNTNGFPATVNDLKAFTEGAVLSFHNICYRVKLSGFLPCRKPVE 70
QY 61 KEILSNINGIMKPGNLAILGPTGGKSLDLVLAARKDPSGLSGDVLINGAPRPNFKCN 120
DB 71 KEILSNINGIMKPGNLAILGPTGGKSLDLVLAARKDPSGLSGDVLINGAPRPNFKCN 130
QY 121 SGYVQDDVVMGTLTVRENLOFSALRLATTMTNHEKNERINRVIEELGLDKVADSKVGT 180
DB 131 SGYVQDDVVMGTLTVRENLOFSALRLATTMTNHEKNERINRVIEELGLDKVADSKVGT 190
QY 181 QIRGVSGGERKRTSIGMELITDPSILSDPTGLDSTANAVALLKRMKSKQGRRTIIF 240
DB 191 QIRGVSGGERKRTSIGMELITDPSILSDPTGLDSTANAVALLKRMKSKQGRRTIIF 250
QY 241 SIHQPRYSIFKLFDSTLLASGRMLPHGPAQALGYFESAGYHCEAYNNPADFFLDIING 300

Db 251 SIHQPRYSIFKLFDSLTLLASGRLMFHGPAQALGYFESAGYHCEAYNNPADFFLDIING 310
QY 301 DSTAVALNREEDFKATEIIEPSKODKPLIEKLAIEIVNSSFYKETHKAEHLQSGGKKKK 360
Db 311 DSTAVALNREEDFKATEIIEPSKODKPLIEKLAIEIVNSSFYKETHKAEHLQSGGKKKK 370
QY 361 ITVFKESITVTSFCHOLRWWSKRSFKNLLGNPOASIAQIIVTVLGLVIGAIYFGLKND 420
Db 371 ITVFKESITVTSFCHOLRWWSKRSFKNLLGNPOASIAQIIVTVLGLVIGAIYFGLKND 430
QY 421 TGIQNRAGVLFLLTNQCFSSVSAVELFVVEKGLFIEHYSIGYRVSSYFLGKLLSDDL 480
Db 431 TGIQNRAGVLFLLTNQCFSSVSAVELFVVEKGLFIEHYSIGYRVSSYFLGKLLSDDL 490
QY 481 MRMLPSIIIFTCIVYFVLMGLKPKADAFVMMFTLMVAYSSASSMALAIAAGQSVSATLL 540
Db 491 MRMLPSIIIFTCIVYFVLMGLKPKADAFVMMFTLMVAYSSASSMALAIAAGQSVSATLL 550
QY 541 MTICFVFMWIFSGLLVNLTTIASWLSWLOYSIPRYGFTALQHNFLQNFPCGLNATGN 600
Db 551 MTICFVFMWIFSGLLVNLTTIASWLSWLOYSIPRYGFTALQHNFLQNFPCGLNATGN 610
QY 601 NPCNATCTGEBYLVKQGDIDSPWGLWKNHVALACMIVIFLTIAVYKLLFLKKYS 655
Db 611 NPCNATCTGEBYLVKQGDIDSPWGLWKNHVALACMIVIFLTIAVYKLLFLKKYS 665

RESULT 13
AAW73627
ID AAW73627 standard; protein; 604 AA.
XX AAW73627;
DT 23-MAR-1999 (first entry)
DE Human secreted protein clone bvl31_5.
KW Secreted protein; human; nutritional supplements; immune stimulant;
KW immune suppressor; haematopoiesis regulator; tissue growth; chemotaxis;
KW activin/inhibin; chemokinesis; haemostasis; thrombolysis;
KW receptor/ligand activity; anti-inflammatory; tumour inhibitor;
KW cadherin/tumour invasion suppressor.
XX Homo sapiens.

Key Location/Qualifiers
FT Misc-difference 85 /note= "unspecified amino acid"
FT FT
XX W09855614-A2.
XX 10-DEC-1998.
XX 01-JUN-1998; 98WO-US11210.
XX 29-MAY-1998; 98US-0087255.
XX 04-JUN-1997; 97US-0868696.
XX 04-JUN-1997; 97US-0868697.
XX 04-JUN-1997; 97US-0868698.
XX 04-JUN-1997; 97US-0868699.
XX 04-JUN-1997; 97US-0868900.
XX 04-JUN-1997; 97US-0869191.
XX 04-JUN-1997; 97US-0869192.
XX 04-JUN-1997; 97US-0869193.
XX 04-JUN-1997; 97US-0869194.
XX (GEM) GENETICS INST INC.
XX Agostino MJ, Fichtel K, Howes SH, Jacobs K, Lavallie ER;
XX McCoy JM, Racie LA, Spaulding V, Treacy M;

DR WPI: 1999-059912/05.
XX N-PSDB; AAV55745.
PT New polynucleotides encoding secreted human proteins - derived from
PT human foetal brain, adult brain, foetal kidney, adult ovary, adult
PT retina, adult placenta or adult uterine cDNA libraries
XX Claim 22; Page 84-86; 127pp; English.
PS This sequence represents a human secreted protein of the invention.
XX The DNA encoding this sequence was isolated from a human adult brain
CC cDNA library, and was designated clone bvl31_5. The DNAs and proteins
CC are predicted to have biological activities which would make them
CC suitable for treating, preventing or ameliorating medical conditions in
CC humans and animals, although no supporting data is given. Suggested
CC activities include nutritional sources or supplements, immune
CC stimulating or suppressing activity, haematopoiesis regulating activity,
CC tissue growth activity, activin/inhibin activity,
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,
CC receptor/ligand activity, anti-inflammatory activity, cadherin/tumour
CC invasion suppressor activity, and tumour inhibition activity. The DNAs
CC are also stated to be useful for gene therapy. A host cell transfected
CC with the DNA, or its subfragments and variants is useful for recombinant
CC production of the human secreted protein clones.
XX SQ Sequence 604 AA;

Query Match 91.0%; Score 3047.5; DB 20; Length 604;
Best Local Similarity 91.8%; Pred. No. 1.le-288;
Matches 601; Conservative 0; Mismatches 3; Indels 51; Gaps 1;
QY 1 MSSNNVEVFIPIVSGQNTNGFPATVSNLDKATEGAVLSFHNICVRVCLKSGFLPCRKPE 60
Db 1 MSSNNVEVFIPIVSGQNTNGFPATVSNLDKATEGAVLSFHNICVRVCLKSGFLPCRKPE 60
QY 61 KEILSNINGIMKPGNLNAILGPTGGKSSLLDLVAARKDPSGLSGDLVINGAPRANFKCN 120
Db 61 KEILSNINGIMKPGNLNAILGPTGGKSSLLDLVAARKDPSGLSGDLVINGAPRANFKCN 120
QY 121 SGYVQDDVVMGTLTVRENLOFSAALRLATWTNHEKNERINRVIEELGDKVADSKVGT 180
Db 121 SGYVQD-----VGT 129
QY 181 QPIRGVSGGERKRTSIGMELITDPSILSLDPTTGLDSTANAVALLKRMKSKQRTIIF 240
Db 130 QPIRGVSGGERKRTSIGMELITDPSILSLDPTTGLDSTANAVALLKRMKSKQRTIIF 189
QY 241 SIHQPRYSIFKLFDSLTLLASGRLMFHGPAQALGYFESAGYHCEAYNNPADFFLDIING 300
Db 190 SIHQPRYSIFKLFDSLTLLASGRLMFHGPAQALGYFESAGYHCEAYNNPADFFLDIING 249
QY 301 DSTAVALNREEDFKATEIIEPSKODKPLIEKLAIEIVNSSFYKETHKAEHLQSGGKKKK 360
Db 250 DSTAVALNREEDFKATEIIEPSKODKPLIEKLAIEIVNSSFYKETHKAEHLQSGGKKKK 309
QY 361 ITVFKESITVTSFCHOLRWWSKRSFKNLLGNPOASIAQIIVTVLGLVIGAIYFGLKND 420
Db 310 ITVFKESITVTSFCHOLRWWSKRSFKNLLGNPOASIAQIIVTVLGLVIGAIYFGLKND 369
QY 421 TGIQNRAGVLFLLTNQCFSSVSAVELFVVEKGLFIEHYSIGYRVSSYFLGKLLSDDL 480
Db 370 TGIQNRAGVLFLLTNQCFSSVSAVELFVVEKGLFIEHYSIGYRVSSYFLGKLLSDDL 429
QY 481 MRMLPSIIIFTCIVYFVLMGLKPKADAFVMMFTLMVAYSSASSMALAIAAGQSVSATLL 540
Db 430 MRMLPSIIIFTCIVYFVLMGLKPKADAFVMMFTLMVAYSSASSMALAIAAGQSVSATLL 489
QY 541 MTICFVFMWIFSGLLVNLTTIASWLSWLOYSIPRYGFTALQHNFLQNFPCGLNATGN 600
Db 490 MTICFVFMWIFSGLLVNLTTIASWLSWLOYSIPRYGFTALQHNFLQNFPCGLNATGN 549
QY 601 NPCNATCTGEBYLVKQGDIDSPWGLWKNHVALACMIVIFLTIAVYKLLFLKKYS 655

Db 550 NPCNYATCTGEBYLVKQIDLSFPMGLWKNHVALACMIVIFLTAYLKLFLKYS 604
ABP61858
ID ABP61858 standard; Protein; 604 AA.
XX
AC ABP61858;
XX
DT 04-OCT-2002 (first entry)
XX
DE Human polypeptide SEQ ID NO 212.
XX
KW Human; cytostatic; antirheumatic; antiarthritic; vulnery; analgesic;
KW antiinflammatory; antibacterial; immunosuppressive; antiparkinsonian;
KW neuroprotective; nootropic; osteopathic; haemostatic; vasotropic;
KW antitumor; fungicide; antidiabetic; antisthmatic; antiallergic;
KW immunostimulant; antiparasitic; secreted protein; transmembrane protein;
KW cytokine; cell proliferation; cell differentiation; autoimmune disease;
KW stem cell; growth factor; nervous system disease; neuropathy;
KW Alzheimer's disease; Parkinson's disease; Huntington's disease;
KW osteoporosis; severe combined immunodeficiency; SCID; infection;
KW multiple sclerosis; rheumatoid arthritis; gene therapy.
XX
OS Homo sapiens.
XX
XX US2002065394-A1.
XX
XX 30-MAY-2002.
XX
XX 22-DEC-2000; 2000US-0745763.
XX
XX 18-MAR-1998; 98US-0040963.
XX
XX (JACO/) JACOBS K.
XX (MCCO/) MCCOY J M.
XX (LAVA/) LAVALLIE E R.
XX (COLL/) COLLINS-RACIE L A.
XX (EVAN/) EVANS C.
XX (MERB/) MERBERG D.
XX (TREA/) TREACY M.
XX (SPAU/) SPAULDING V.
XX
XX Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
XX Merberg D, Treacy M, Spaulding V;
XX
XX WPI: 2002-582343/62.
XX N-PSDB; ABQ92071.
XX
XX Novel secreted or transmembrane protein and polynucleotide encoding the
XX protein, useful for diagnosis and treatment of neurological disorders,
XX cancer, autoimmune diseases, bone disorders and lung or liver fibrosis
XX
XX Claim 248; Page 229-231; 284pp; English.
XX
XX The invention relates to human secreted or transmembrane protein (I),
XX their fragments and is encoded by specific complementary deoxyribonucleic
XX acid (cDNA) inserts (II), where the protein is substantially free from
XX other mammalian proteins. (I) are useful for preventing, treating or
XX ameliorating a medical condition, especially immunological treatment or
XX prevention of tumours. (I) exhibits activity relating to angiogenesis,
XX cytokine, cell proliferation, cell differentiation, antiinflammatory,
XX stem cell growth factor activity and activin or inhibin-related
XX activities. (I) can be used to manipulate stem cells in culture to give
XX rise to neuroepithelial cells that can be used to augment or replace
XX cells damaged by illness, autoimmune disease, accidental damage or
XX genetic disorders. (I) induces the proliferation of neural cells and
XX regeneration of nerve and brain tissue and is useful for the treatment of
XX central and peripheral nervous system diseases and neuropathies, such as
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX lateral sclerosis. (I) is involved in chemotactic or chemokinetic
XX activity, regulation of haematopoiesis and is useful for treating myeloid

CC or lymphoid cell disorders, platelet disorders such as thrombocytopaenia
CC and for regeneration of bone, cartilage, tendon, ligament and/or nerve
CC tissue growth and in tissue repair, healing of burns, incisions, ulcers,
CC for treating osteoporosis, osteoarthritis, bone degenerative disorders or
CC periodontal disease. (I) is also useful for gut protection or
CC regeneration and treatment of lung or liver fibrosis, reperfusion injury
CC in various tissues, various immune deficiencies and disorders including
CC severe combined immunodeficiency (SCID), bacterial or fungal infections,
CC autoimmune disorders e.g. multiple sclerosis, rheumatoid arthritis,
CC diabetes mellitus, myasthenia gravis, allergic reactions and conditions,
CC such as asthma or other respiratory problems. (II) is useful to express
CC recombinant protein, as markers for tissues in which the corresponding
CC protein is preferentially expressed and in gene therapy. The present
CC sequence is that of a polypeptide of the invention.
XX
SQ Sequence 604 AA;
Query Match 91.0%; Score 3047.5; DB 23; Length 604;
Best Local Similarity 91.8%; Pred. NO. 1.le-289;
Matches 601; Conservative 0; Mismatches 3; Indels 51; Gaps 1;
QY 1 MSSNVVEFIPVSCQNTNGPPATVSNDLKAFTEGAVLSFHNICVRVVKLSGFLPCRKPV 60
DB 1 MSSNVVEFIPVSCQNTNGPPATVSNDLKAFTEGAVLSFHNICVRVVKLSGFLPCRKPV 60
QY 61 KEILSNINGIMKPGINAILGPTGGKSSLLDLAARKDPSGLSDVLINGAPRANPKCN 120
DB 61 KEILSNINGIMKPGINAILGPTGGKSSLLDLAARKDPSGLSDVLINGAPRANPKCN 120
QY 121 SGYVQDDVVMGTLTVRENLFQSAALRLATWTNHEKNERINRVIEELGDKVADSKVGT 180
DB 121 SGYVQ-----VGT 129
QY 181 QFIRGSGGERKRTSIGMELITDPSILSLDEPTTGLDSDSTANAVALLKRMKSKQRTIIP 240
DB 130 QFIRGSGGERKRTSIGMELITDPSILSLDEPTTGLDSDSTANAVALLKRMKSKQRTIIP 189
QY 241 SIHQPRYSIFKLFDSLTLTLLASGRMLFHGPAOEALGYFESAGYHCEAYNNPADFFLDIING 300
DB 190 SIHQPRYSIFKLFDSLTLTLLASGRMLFHGPAOEALGYFESAGYHCEAYNNPADFFLDIING 249
QY 301 DSTAVALNREDFKATEIIEPSKQDKPLIEKLAIEIYVNSFPYKTKABELHQLSGGEKKKK 360
DB 250 DSTAVALNREDFKATEIIEPSKQDKPLIEKLAIEIYVNSFPYKTKABELHQLSGGEKKKK 309
QY 361 ITVFKIEISYTTSFCHOLRWVSKRSPKNLLGNPQASIAQIIVTVVLGLVIGAIYFGLKNDS 420
DB 310 ITVFKIEISYTTSFCHOLRWVSKRSPKNLLGNPQASIAQIIVTVVLGLVIGAIYFGLKNDS 369
QY 421 TGIQNRAGVLPFLTITNOCFSSVSAVELFVVEKKLFIHEYISGYRVSSYFGLKLSDL 480
DB 370 TGIQNRAGVLPFLTITNOCFSSVSAVELFVVEKKLFIHEYISGYRVSSYFGLKLSDL 429
QY 481 NRMPLPSIIFTCIVYFMLGLKPKADAPFVMMFTLMMVAYSASSMALAIAAGQSVSVATLL 540
DB 430 NRMPLPSIIFTCIVYFMLGLKPKADAPFVMMFTLMMVAYSASSMALAIAAGQSVSVATLL 489
QY 541 MTICFVPMWIPSGLLVNLTTTASLWSLQYFSPRYGFTALOHNEFLQNCPCPNATGN 600
DB 490 MTICFVPMWIPSGLLVNLTTTASLWSLQYFSPRYGFTALOHNEFLQNCPCPNATGN 549
QY 601 NPCNYATCTGEBYLVKQIDLSFPMGLWKNHVALACMIVIFLTAYLKLFLKYS 655
DB 550 NPCNYATCTGEBYLVKQIDLSFPMGLWKNHVALACMIVIFLTAYLKLFLKYS 604
RESULT 15
ID ABB07272 standard; Protein; 657 AA.
XX
AC ABB07272;
XX
DT 26-MAR-2002 (first entry)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 1, 2003, 18:38:06 ; Search time 22.3616 Seconds
(without alignments)
1043.432 Million cell updates/sec

Title: US-09-856-927-4
Perfect score: 769
Sequence: 1 FGLGAEYATSSMALAIATG.....MIIIFLTATYKLFLKYS 147

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_13Jun03.*
1: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-embl/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-embl/AA2001.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-embl/AA2002.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-embl/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	769	100.0	147 21	AA1995366 ATP-binding caset
2	769	100.0	147 23	ABB07271 Mouse BCRP (mBCRP)
3	739	96.1	657 22	ABB07272 Murine BCRP (mBCRP)
4	623	81.0	456 22	AA1993564 Human protein sequ
5	623	81.0	604 20	AA1993564 Human secreted pro
6	623	81.0	604 23	ABP61858 Human polypeptide
7	623	81.0	655 21	AA1993565 ATP-binding caset
8	623	81.0	655 22	AA1993565 Human BCRP/MXR/ABC
9	623	81.0	655 22	AA1993565 Human transport pr

10	623	81.0	655	23	ABP52127	Homo sapiens ABC t
11	623	81.0	655	23	AAU80028	Human ABCG2. Homo
12	623	81.0	655	23	AAU80029	Human ABCG2 mutant
13	623	81.0	655	23	AAO14781	Human BCRP protein
14	623	81.0	655	23	ABB07270	Human BCRP (huBCRP)
15	623	81.0	655	23	ABB07273	Human BCRP (huBCRP)
16	623	81.0	663	20	AAI15221	Breast Cancer Resi
17	616	80.1	665	23	AAO14782	Human BCRP-related
18	616	80.1	665	23	AAO14783	Human BCRP-related
19	258	33.6	73	23	ABG93686	Human Mitoxantone
20	243.5	31.7	72	23	ABG93710	Human Mitoxantone
21	163	21.2	41	23	ABG93687	Human Mitoxantone
22	163	21.2	41	23	ABG93731	Human Mitoxantone
23	128	16.6	687	22	ABBS9384	Drosophila melanog
24	126.5	16.4	646	23	AAE28968	Human ABCG4 transp
25	125.5	16.3	687	21	AAV78981	Silkworm Bm white
26	124.5	16.2	1499	23	ABP52138	Candida albicans m
27	123.5	16.1	646	23	ABBS98348	Human ABC transpor
28	123.5	16.1	646	23	AAE28964	Human ABCG4 transp
29	123.5	16.1	646	23	AAO14186	Human transporter
30	123.5	16.1	646	24	ABBS9694	Amino acid sequenc
31	123.5	16.1	646	24	ABBS9696	Amino acid sequenc
32	122.5	15.9	935	22	AA178589	Human protein SEQ
33	122.5	15.9	935	22	AA178589	Human bone marrow
34	117	15.2	674	23	ABP52126	Homo sapiens ABC t
35	113	14.7	638	23	ABBS98349	Human ABC transpor
36	113	14.7	663	24	ABBS82647	Human DevG22 homol
37	110	14.3	666	23	ABBS7112	Mouse ischaemic co
38	109.5	14.2	1501	23	ABP52137	Candida albicans m
39	109.5	14.2	1501	23	ABP52140	Saccharomyces cere
40	108	14.0	24	22	AAU04318	ATP-binding caset
41	103.5	13.5	620	22	ABBS9648	Drosophila melanog
42	102.5	13.3	689	21	AAV78980	Silkworm Bm white
43	97.5	12.7	689	22	ABBS1867	Drosophila melanog
44	97.5	12.7	832	22	ABBS9544	Drosophila melanog
45	94	12.2	19	23	ABG93689	Human Mitoxantone

ALIGNMENTS

RESULT 1
AA1995366
ID AA1995366 standard; Protein; 147 AA.
AC AA1995366;
XX
DT 25-SEP-2000 (first entry)
DE ATP-binding cassette protein MXR2 C-terminal region.
KW ATP-binding cassette protein; ABC protein; MXR2; human;
KW cytotocin resistance; colon carcinoma; therapy.
XX
OS Homo sapiens.
XX
PN WO200036101-A2.
XX
PD 22-JUN-2000.
XX
PF 24-NOV-1999; 99WO-US28107.
XX
PR 30-NOV-1998; 98US-0110473.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Dean M, Allikmets R, Bates SE, Fojo AT;
XX
DR WPI; 2000-442385/38.
XX
PT N-PSDB; AAA27939.
XX
PT Adenosine triphosphate (ATP)-binding cassette proteins and their
XX nucleic acids, useful for identifying agents that may be used to treat

PT colonic carcinomas resistant to cytotoxic therapeutic agents -
XX
PS Claim 1(ii); Page 46; 49pp; English.
XX

CC The present sequence is that of the C-terminal region of an
CC ATP-binding cassette (ABC) protein designated MXR2 that confers
CC mitoxantrone resistance to S1-M1-80 human colon carcinoma cells.
CC Nucleic acids encoding MXR2 may be used to produce the protein
CC using recombinant DNA methodologies. The expressed proteins may
CC be used to confer resistance to cytotoxins such as mitoxantrone
CC and/or daunomycin to the host cells in which they are expressed.
CC In disease conditions such as carcinoma of the colon, breast and
CC gastrointestinal tract, the ABC protein may be expressed to provide
CC resistance to cytotoxic therapeutic agents. The nucleic acids and
CC proteins may be used to produce host cell models of resistant cells
CC which can be used to screen for candidate agents that inhibit the
CC expression and/or activity of the ABC protein. These agents may be
CC used to down regulate ABC protein expression in carcinoma cells and
CC sensitize them to cytotoxic therapeutic agents.

SQ Sequence 147 AA;

Query Match 100.0%; Score 769; DB 21; Length 147;
Best Local Similarity 100.0%; Pred. No. 6.7e-79;
Matches 147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FGLGAEAYTASSMALAIATGQSVSVATLLMTIAFVFMFLFSGLLVNLRTIGPWLWLOQ 60
DB 1 FGLGAEAYTASSMALAIATGQSVSVATLLMTIAFVFMFLFSGLLVNLRTIGPWLWLOQ 60
QY 61 FSPRYGFTALQYNEFLGQFCPGFNVTNDSTCVNSYAICTGNEYLINOGIELSPWGLWK 120
DB 61 FSPRYGFTALQYNEFLGQFCPGFNVTNDSTCVNSYAICTGNEYLINOGIELSPWGLWK 120
QY 121 NHVALACMIIFLTIAYLKLLFLKKYS 147
DB 121 NHVALACMIIFLTIAYLKLLFLKKYS 147

RESULT 2

ABB07271
ID ABB07271 standard; Protein; 147 AA.

XX AC ABB07271;

XX DT 26-MAR-2002 (first entry)

XX DE Mouse BCRP (mBCRP) amino acid sequence.

XX KW Stem cell; ATP transport protein; ATP-binding cassette; antiparkinsonian;
XX KW hepatotropic; neurodegenerative; cytostatic; antianemic; muscular; BCRP;
XX KW cardiant; gene therapy.

XX OS Mus musculus.

XX PN WO200192877-A2.

XX PD 06-DEC-2001.

XX PF 30-MAY-2001; 2001WO-US17459.

XX PP 31-MAY-2000; 2000US-0584586.

XX PR 29-MAY-2001; 2001US-0866866.

XX PA (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.

XX PI Sorrentino B, Schuetz J;

XX DR WPI; 2002-114368/15.

XX DR N-PSDB; ABA94370.

XX Identifying a stem cell, for treating e.g., muscular dystrophy,
PT myocardial infarction, Parkinson's disease, or neurodegenerative

PT disorders, comprises detecting the expression of an ATP transport
PT protein (BCRP) by a cell -
XX
PS Examples; Page 80; 87pp; English.

CC The invention provides a method of identifying and/or isolating a stem
CC cell that involves detecting the expression of an ATP transport protein
CC containing a conserved ATP-binding cassette (BCRP) by a cell in a sample
CC comprising stem cells. The isolated stem cells may be used in the
CC treatment of diseases such as muscular dystrophy, degenerative liver
CC disorder, myocardial infarction, Parkinson's disease, degenerative
CC disorders of the brain, and for tissue regeneration or replacement.
CC Haematopoietic cells can be used in bone marrow transplants (e.g., for
CC treatment of leukemia) and for ex vivo gene therapy for treating blood
CC diseases such as sickle cell anemia and thalassemia. The stem cells can
CC also be used as cell targets in gene therapy protocols. The present
CC sequence represents the amino acid sequence of mouse BCRP.

SQ Sequence 147 AA;

Query Match 100.0%; Score 769; DB 23; Length 147;
Best Local Similarity 100.0%; Pred. No. 6.7e-79;
Matches 147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FGLGAEAYTASSMALAIATGQSVSVATLLMTIAFVFMFLFSGLLVNLRTIGPWLWLOQ 60
DB 1 FGLGAEAYTASSMALAIATGQSVSVATLLMTIAFVFMFLFSGLLVNLRTIGPWLWLOQ 60
QY 61 FSPRYGFTALQYNEFLGQFCPGFNVTNDSTCVNSYAICTGNEYLINOGIELSPWGLWK 120
DB 61 FSPRYGFTALQYNEFLGQFCPGFNVTNDSTCVNSYAICTGNEYLINOGIELSPWGLWK 120
QY 121 NHVALACMIIFLTIAYLKLLFLKKYS 147
DB 121 NHVALACMIIFLTIAYLKLLFLKKYS 147

RESULT 3

ABB07272
ID ABB07272 standard; Protein; 657 AA.

XX AC ABB07272;

XX DT 26-MAR-2002 (first entry)

XX DE Murine BCRP (mBCRP) amino acid sequence.

XX KW Stem cell; ATP transport protein; ATP-binding cassette; antiparkinsonian;
XX KW hepatotropic; neurodegenerative; cytostatic; antianemic; muscular; BCRP;
XX KW cardiant; gene therapy.

XX OS Mus musculus.

XX PN WO200192877-A2.

XX PD 06-DEC-2001.

XX PF 30-MAY-2001; 2001WO-US17459.

XX PP 31-MAY-2000; 2000US-0584586.

XX PR 29-MAY-2001; 2001US-0866866.

XX PA (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.

XX PI Sorrentino B, Schuetz J;

XX DR WPI; 2002-114368/15.

XX DR N-PSDB; ABA94371.

XX Identifying a stem cell, for treating e.g., muscular dystrophy,
PT myocardial infarction, Parkinson's disease, or neurodegenerative
PT disorders, comprises detecting the expression of an ATP transport
PT protein (BCRP) by a cell -

```
XX Examples; Page 82-83; 87pp; English.
PS The invention provides a method of identifying and/or isolating a stem
CC cell that involves detecting the expression of an ATP transport protein
CC containing a conserved ATP-binding cassette (BCRP) by a cell in a sample
CC comprising stem cells. The isolated stem cells may be used in the
CC treatment of diseases such as muscular dystrophy, degenerative liver
CC disorder, myocardial infarction, Parkinson's disease, degenerative
CC disorders of the brain, and for tissue regeneration or replacement.
CC Haematopoietic cells can be used in bone marrow transplants (e.g., for
CC treatment of leukemia) and for ex vivo gene therapy for treating blood
CC diseases such as sickle cell anemia and thalassemia. The stem cells can
CC also be used as cell targets in gene therapy protocols. The present
CC sequence represents the amino acid sequence of mouse BCRP.
XX
SQ Sequence 657 AA;

Query Match 96.1%; Score 739; DB 23; Length 657;
Best Local Similarity 97.3%; Pred. No. 1.2e-74;
Matches 143; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 FGLGAEAYTASSMALAIATGQSVSVATLLMTIAFVPMFLFSGLLVNLRTIGPWLSQLQY 60
DB 511 FTLIMVAYTASSMALAIATGQSVSVATLLMTIAFVPMFLFSGLLVNLRTIGPWLSQLQY 570

QY 61 FSIPRYGFTALQYNEFLQGFCEPGFNVTNSTCVNSYAICTGNEVYLINQGIELSPWGLWK 120
DB 571 FSIPRYGFTALQYNEFLQGFCEPGFNVTNSTCVNSYAICTGNEVYLINQGIELSPWGLWK 630

QY 121 NHVALACMIIFLTAYLKLFLKKYS 147
DB 631 NHVALACMIIFLTAYLKLFLKKYS 657

RESULT 4
AAB93564
ID AAB93564 standard; Protein; 456 AA.
AC AAB93564;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:12962.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
XX
PR 27-AUG-1999; 99JP-0300253.
XX
PR 11-JAN-2000; 2000JP-0118776.
XX
PR 02-MAY-2000; 2000JP-0183767.
XX
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayaashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PR Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
```

```
PS Claim 8; SEQ ID 12962; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 456 AA;

Query Match 81.0%; Score 623; DB 22; Length 456;
Best Local Similarity 81.6%; Pred. No. 9.6e-62;
Matches 120; Conservative 8; Mismatches 17; Indels 2; Gaps 1;

QY 1 FGLGAEAYTASSMALAIATGQSVSVATLLMTIAFVPMFLFSGLLVNLRTIGPWLSQLQY 60
DB 312 FTLIMVAYTASSMALAIATGQSVSVATLLMTIAFVPMFLFSGLLVNLRTIGPWLSQLQY 371

QY 61 FSIPRYGFTALQYNEFLQGFCEPGFNVTNSTCVNSYAICTGNEVYLINQGIELSPWGLWK 120
DB 372 FSIPRYGFTALQYNEFLQGFCEPGFNVTNSTCVNSYAICTGNEVYLINQGIELSPWGLWK 429

QY 121 NHVALACMIIFLTAYLKLFLKKYS 147
DB 430 NHVALACMIIFLTAYLKLFLKKYS 456

RESULT 5
AAW73627
ID AAW73627 standard; protein; 604 AA.
XX
AC AAW73627;
XX
DT 23-MAR-1999 (first entry)
XX
DE Human secreted protein clone bv131_5.
XX
KW Secreted protein; human; nutritional supplements; immune stimulant;
KW immune suppressor; haematopoiesis regulator; tissue growth; chemotaxis;
KW activin/inhibin; chemokinesis; haemostasis; thrombolytic;
KW receptor/ligand activity; anti-inflammatory; tumour inhibitor;
KW cadherin/tumour invasion suppressor.
XX
OS Homo sapiens.
XX
PH Key Location/Qualifiers
FT Misc-difference 85 /note= "unspecified amino acid"
XX
PN WO9855614-A2.
XX
PD 10-DEC-1998.
XX
PF 01-JUN-1998; 98WO-US11210.
XX
PR 29-MAY-1998; 98US-0087255.
XX
```

PR 04-JUN-1997; 97US-0868696.
 PR 04-JUN-1997; 97US-0868697.
 PR 04-JUN-1997; 97US-0868698.
 PR 04-JUN-1997; 97US-0868699.
 PR 04-JUN-1997; 97US-0868900.
 PR 04-JUN-1997; 97US-0869191.
 PR 04-JUN-1997; 97US-0869192.
 PR 04-JUN-1997; 97US-0869193.
 PR 04-JUN-1997; 97US-0869194.
 XX (GEM) GENETICS INST INC.
 XX
 XX Agostino MJ, Fecthel K, Howes SH, Jacobs K, Lavallie ER;
 PI McCoy JM, Racie LA, Spaulding V, Treacy M;
 XX WPI; 1999-059912/05.
 DR N-PSDB; AAV55745.
 XX
 XX New polynucleotides encoding secreted human proteins - derived from
 PT human foetal brain, adult brain, foetal kidney, adult ovary, adult
 PT retina, adult placenta or adult uterus cDNA libraries
 XX
 XX Claim 22; Page 84-86; 127pp; English.
 XX This sequence represents a human secreted protein of the invention.
 CC The DNA encoding this sequence was isolated from a human adult brain
 CC cDNA library, and was designated clone bvi31_5. The DNAs and proteins
 CC are predicted to have biological activities which would make them
 CC suitable for treating, preventing or ameliorating medical conditions in
 CC humans and animals, although no supporting data is given. Suggested
 CC activities include nutritional sources or supplements, immune
 CC stimulating or suppressing activity, haematopoiesis regulating activity,
 CC tissue growth activity, activin/inhibin activity,
 CC chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,
 CC receptor/ligand activity, anti-inflammatory activity, cadherin/tumour
 CC invasion suppressor activity, and tumour inhibition activity. The DNAs
 CC are also stated to be useful for gene therapy. A host cell transfected
 CC with the DNA, or its subfragments and variants is useful for recombinant
 CC production of the human secreted protein clones.
 XX Sequence 604 AA;
 XX
 XX Query Match 81.0%; Score 623; DB 20; Length 604;
 XX Best Local Similarity 81.6%; Pred. No. 1.4e-61;
 XX Matches 120; Conservative 8; Mismatches 17; Indels 2; Gaps 1;
 QY 1 FGLGAEATPASSMALAIATGQSVSVATLLMTIAFVFMFLFSGLLVNLRTIGFWLSWLQY 60
 Db 460 FTLMVAVSASSMALAIAGQSVSVATLLMTICFVFMIFSGLLVNLRTIASWLSWLQY 519
 QY 61 FSPRYGFTALQYNFLQGFPGFVNTDSTCVNSYAICTGNEYLINQIGIELSPWGLWK 120
 Db 520 FSPRYGFTALQHNFLQGFPGFGLNATGNPC--NYATCTGEEYLVKQIGDLSWGLWK 577
 QY 121 NHVALACMIITETAYLKLLFLKKYS 147
 Db 578 NHVALACMIIVETAYLKLLFLKKYS 604
 RESULT 6
 ABP61858
 ID ABP61858 standard; Protein; 604 AA.
 XX
 AC ABP61858;
 XX
 DT 04-OCT-2002 (first entry)
 XX
 DE Human polypeptide SEQ ID NO 212.
 XX
 XX Human; cytostatic; antirheumatic; antiarthritic; vulnery; analgesic;
 KW antinflammatory; antibacterial; immunosuppressive; antiparkinsonian;
 KW neuroprotective; nootropic; osteopathic; haemostatic; vasotropic;

KW antiulcer; fungicide; antidiabetic; antiasthmatic; antiallergic;
 KW immunostimulant; antiparasitic; secreted protein; transmembrane protein;
 KW cytokine; cell proliferation; cell differentiation; autoimmune disease;
 KW stem cell; growth factor; nervous system disease; neuropathy;
 KW Alzheimer's disease; Parkinson's disease; Huntington's disease;
 KW osteoporosis; severe combined immunodeficiency; SCID; infection;
 KW multiple sclerosis; rheumatoid arthritis; gene therapy.
 XX Homo sapiens.
 XX US2002065394-A1.
 XX 30-MAY-2002.
 XX 22-DEC-2000; 2000US-0745763.
 XX 18-MAR-1998; 98US-0040963.
 XX (JACO/) JACOBS K.
 XX (MCCO/) MCCOY J M.
 XX (LAVA/) LAVALLIE E R.
 XX (COLL/) COLLINS-RACIE L A.
 XX (EVAN/) EVANS C.
 XX (MERB/) MERBERG D.
 XX (TREAC/) TREACY M.
 XX (SPAU/) SPAULDING V.
 XX Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
 PI Merberg D, Treacy M, Spaulding V;
 XX WPI; 2002-582343/62.
 DR N-PSDB; ABQ92071.
 XX Novel secreted or transmembrane protein and polynucleotide encoding the
 PT protein, useful for diagnosis and treatment of neurological disorders,
 PT cancer, autoimmune diseases, bone disorders and lung or liver fibrosis
 PT
 XX Claim 248; Page 229-231; 284pp; English.
 CC The invention relates to human secreted or transmembrane protein (I),
 CC their fragments and is encoded by specific complementary deoxyribonucleic
 CC acid (cDNA) inserts (II), where the protein is substantially free from
 CC other mammalian proteins. (I) are useful for preventing, treating or
 CC ameliorating a medical condition, especially immunological treatment or
 CC prevention of tumours. (I) exhibits activity relating to angiogenesis,
 CC cytokine, cell proliferation, cell differentiation, anti-inflammatory,
 CC stem cell growth factor activity and activin or inhibin-related
 CC activities. (I) can be used to manipulate stem cells in culture to give
 CC rise to neuroepithelial cells that can be used to augment or replace
 CC cells damaged by illness, autoimmune disease, accidental damage or
 CC genetic disorders. (I) induces the proliferation of neural cells and
 CC regeneration of nerve and brain tissue and is useful for the treatment of
 CC central and peripheral nervous system diseases and neuropathies, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis. (I) is involved in chemotactic or chemokinetic
 CC activity, regulation of haematopoiesis and is useful for treating myeloid
 CC or lymphoid cell disorders, platelet disorders such as thrombocytopenia
 CC and for regeneration of bone, cartilage, tendon, ligament and/or nerve
 CC tissue growth and in tissue repair, healing of burns, incisions, ulcers,
 CC for treating osteoporosis, osteoarthritis, bone degenerative disorders or
 CC periodontal disease. (I) is also useful for gut protection or
 CC regeneration and treatment of lung or liver fibrosis, reperfusion injury
 CC in various tissues, various immune deficiencies and disorders including
 CC severe combined immunodeficiency (SCID), bacterial or fungal infections,
 CC autoimmune disorders e.g. multiple sclerosis, rheumatoid arthritis,
 CC diabetes mellitus, myasthenia gravis, allergic reactions and conditions,
 CC such as asthma or other respiratory problems. (II) is useful to express
 CC recombinant protein, as markers for tissues in which the corresponding
 CC protein is preferentially expressed and in gene therapy. The present
 CC sequence is that of a polypeptide of the invention.
 XX Sequence 604 AA;

Query Match 81.0%; Score 623; DB 23; Length 604;
Best Local Similarity 81.6%; Pred. NO. 1.4e-61;
Matches 120; Conservative 8; Mismatches 17; Indels 2; Gaps 1;

QY 1 FGLGAEATASSMALAIATGOSVSVATLLMTIAFVFMFLFSGLLVNLRTTIGPWSLWQY 60
DB 460 FTLMMVAYSASSMALAIAGOSVSVATLLMTICFVFMIFSGLLVNLRTTIASWLSWQY 519
QY 61 FSPRYGFTALQVNEFLQGEFCPGFNVTNSTCVNSYAICTGNEYLINQGIELSPLWGLWK 120
DB 520 FSPRYGFTALQVNEFLQGEFCPGFNVTNSTCVNSYAICTGNEYLINQGIELSPLWGLWK 120
QY 121 NHVALACMIIFLTIAYLKLLFLKKYS 147
DB 578 NHVALACMIIFLTIAYLKLLFLKKYS 604

RESULT 7
AAY95365
ID AAY95365 standard; Protein; 655 AA.
XX
AC AAY95365;
XX
DT 25-SEP-2000 (first entry)
XX
DE ATP-binding cassette protein MXR1.
XX
KW ATP-binding cassette protein; ABC protein; MXR1; human;
KW cytotoxin resistance; colon carcinoma; therapy.
XX
OS Homo sapiens.
XX
PN WO200036101-A2.
XX
PD 22-JUN-2000.
XX
PF 24-NOV-1999; 99WO-US28107.
XX
PR 30-NOV-1999; 98US-0110473.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Dean M, Allikmets R, Bates SE, Fojo AT;
XX
DR WPI; 2000-442385/38.
XX
DR N-PSDB; AA27938.
XX
XX Adenosine triphosphate (ATP)-binding cassette proteins and their
PT nucleic acids, useful for identifying agents that may be used to treat
PT colonic carcinomas resistant to cytotoxic therapeutic agents -
XX
PS Claim 1(ii); Page 46; 49pp; English.
XX
CC The present sequence is that of an ATP-binding cassette protein
CC (ABC protein) designated MXR1 that confers mitoxantrone resistance
CC to S1-M1-80 human colon carcinoma cells. Nucleic acids encoding
CC the ABC protein may be used to produce the protein according to
CC standard recombinant DNA methodologies. The expressed proteins may
CC be used to confer resistance to cytotoxins such as mitoxantrone
CC and/or daunomycin to the host cells in which they are expressed.
CC In disease conditions such as carcinoma of the colon, breast and
CC gastrointestinal tract, the ABC protein may be expressed to provide
CC resistance to cytotoxic therapeutic agents. The nucleic acids and
CC proteins may be used to produce host cell models of resistant cells
CC which can be used to screen for candidate agents that inhibit the
CC expression and/or activity of the ABC protein. These agents may be
CC used to down regulate ABC protein expression in carcinoma cells and
CC sensitize them to cytotoxic therapeutic agents.

SQ Sequence 655 AA;

Query Match

81.0%; Score 623; DB 21; Length 655;

Best Local Similarity 81.6%; Pred. NO. 1.5e-61;
Matches 120; Conservative 8; Mismatches 17; Indels 2; Gaps 1;

QY 1 FGLGAEATASSMALAIATGOSVSVATLLMTIAFVFMFLFSGLLVNLRTTIGPWSLWQY 60
DB 511 FTLMMVAYSASSMALAIAGOSVSVATLLMTICFVFMIFSGLLVNLRTTIASWLSWQY 570
QY 61 FSPRYGFTALQVNEFLQGEFCPGFNVTNSTCVNSYAICTGNEYLINQGIELSPLWGLWK 120
DB 571 FSPRYGFTALQVNEFLQGEFCPGFNVTNSTCVNSYAICTGNEYLINQGIELSPLWGLWK 628
QY 121 NHVALACMIIFLTIAYLKLLFLKKYS 147
DB 629 NHVALACMIIFLTIAYLKLLFLKKYS 655

RESULT 8
AAU04348
ID AAU04348 standard; Protein; 655 AA.
XX
AC AAU04348;
XX
DT 23-OCT-2001 (first entry)
XX
DE Human BCRP/MXR/ABCP protein.
XX
KW ATP-binding cassette transporter protein-inhibiting peptide;
KW ABC; transmembrane domain; cancer; tumour; HIV; AIDS; BCRP/MXR/ABCP;
KW acquired immunodeficiency syndrome; human immunodeficiency syndrome.
XX
OS Homo sapiens.
XX
FH Key
FT Domain 398..420
FT /label= Transmembrane_domain_1
FT Domain 427..450
FT /label= Transmembrane_domain_2
FT Domain 478..501
FT /label= Transmembrane_domain_3
FT Domain 504..527
FT /label= Transmembrane_domain_4
FT Domain 539..558
FT /label= Transmembrane_domain_5
FT Domain 629..650
FT /label= Transmembrane_domain_6
XX
PN WO200136477-A2.
XX
PD 25-MAY-2001.
XX
PF 17-NOV-2000; 2000WO-US31817.
XX
PR 18-NOV-1999; 99US-0166382.
PR 22-NOV-1999; 99US-0166767.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Tarasova NI, Michejda CJ, Gottesman MM, Hrycyna C;
XX
DR WPI; 2001-381224/40.
XX
CC New ATP-binding cassette transporter-inhibiting peptides derived from
CC transmembrane domain of the transporter, useful for inhibiting
CC biological activity of the transporter and sensitizing cancer cells to
CC chemotherapeutics -
XX
PS Disclosure; Fig 2; 89pp; English.
XX
CC The sequence represents human BCRP/MXR/ABCP protein, an ATP-binding
CC cassette (ABC) transporter protein. Peptides derived from its
CC transmembrane domains are used as ATP-binding cassette
CC transporter protein-inhibiting peptides. The peptides have a first end
CC and a second end, where at the first end there is a group that is

CC negatively charged under physiological conditions and at the second end a
 CC group that is neutrally charged under physiological conditions and is at
 CC least 70% identical to a transmembrane domain of an ABC transporter. The
 CC peptides are used for inhibiting the biological activity, in particular
 CC ion flux or translocation, cytochrome efflux or translocation,
 CC phosphorylation, protein synthesis or degradation, cellular morphology,
 CC secretion, production of particular components such as soluble
 CC inositol phosphates, tumour growth, chemotaxis, mitogenic response, cell
 CC growth activation or secretion of a target ATP-binding cassette (ABC)
 CC transporter. The peptides are useful to sensitize cancer
 CC cells to standard chemotherapeutics, which chemotherapeutics are then
 CC administered to kill the cancer cells. Transmembrane analogues of ABC
 CC transporter protein are administered to patients infected with HIV-1,
 CC who are taking a protease inhibitor. The ability of HIV-1 infected
 CC cells to efflux the protease inhibitor is impaired, resulting in higher
 CC concentration of the protease inhibitor in the affected cell. This
 CC reduces the amount of protease inhibitor administered to the patient to
 CC achieve a virus-inhibiting dose in the infected cells.

XX Sequence 655 AA;

Query Match 81.0%; Score 623; DB 22; Length 655;

Best Local Similarity 81.6%; Pred. No. 1.5e-61;

Matches 120; Conservative 8; Mismatches 17; Indels 2; Gaps 1;

QY 1 FGLGAEAYTASSMALAIATGQSVSVATLLMTIAFVFMFLFGLLVNLTIGPWLWQY 60

DB 511 FTLMMVAYSASSMALAIAGQSVSVATLLMTICFVFMIFSGLLVNLTTIASWLSWQY 570

QY 61 FSIPIRYGFTALQYNEFLGQEFPCGFNVTDNSTCVNSYAICTGNEYLINQGIELSPWGLWK 120

DB 571 FSIPIRYGFTALQYNEFLGQEFPCGFNLATGNNPC--NYATCTGEYLVKQIDLSWGLWK 628

QY 121 NHVALACMIIFLTIAYLKLLFLKKYS 147

DB 629 NHVALACMIVIFLTIAYLKLLFLKKYS 655

RESULT 9

AAB60104

ID AAB60104 standard; Protein; 655 AA.

AC AAB60104;

DT 28-MAR-2001 (first entry)

XX Human transport protein TPPT-24.

XX Human; transport protein; TPPT; transport disorder; metabolic disorder;
 KW neurological disorder; cardiovascular disorder; reproductive disorder;
 KW immune disorder; cancer.

OS Homo sapiens.

PN WO200078953-A2.

XX 28-DEC-2000.

PF 16-JUN-2000; 2000WO-US16668.

XX 17-JUN-1999; 99US-0139923.

PR 10-AUG-1999; 99US-0148177.

PR 18-AUG-1999; 99US-0149357.

PR 28-OCT-1999; 99US-0162287.

XX (INCY-) INCYTE GENOMICS INC.

XX Lal P, Yang J, Yue H, Hillman JL, Tang YT, Bandman O, Burford N;

PI Baughn MR, Azimzai Y, Lu DAM, Au-Young J, Patterson C;

XX WPI; 2001-041424/05.

DR N-PSDB; AAF27724.

XX

PT Isolated polypeptide with a human transport protein sequence is useful
 PT for the diagnosis, prevention and treatment of disorders associated
 PT with the immune, reproductive and cardiovascular systems -

XX Claim 2; Page 126-127; 165pp; English.

XX The present invention provides the protein and coding sequences for 43
 CC novel human transport proteins (designated TPPTs). These can be used in
 CC the diagnosis and treatment of transport, metabolic, neurological,
 CC reproductive, cardiovascular and immune disorders, and cell proliferative
 CC disorders such as cancer.

XX Sequence 655 AA;

Query Match 81.0%; Score 623; DB 22; Length 655;

Best Local Similarity 81.6%; Pred. No. 1.5e-61;

Matches 120; Conservative 8; Mismatches 17; Indels 2; Gaps 1;

QY 1 FGLGAEAYTASSMALAIATGQSVSVATLLMTIAFVFMFLFGLLVNLTIGPWLWQY 60

DB 511 FTLMMVAYSASSMALAIAGQSVSVATLLMTICFVFMIFSGLLVNLTTIASWLSWQY 570

QY 61 FSIPIRYGFTALQYNEFLGQEFPCGFNVTDNSTCVNSYAICTGNEYLINQGIELSPWGLWK 120

DB 571 FSIPIRYGFTALQYNEFLGQEFPCGFNLATGNNPC--NYATCTGEYLVKQIDLSWGLWK 628

QY 121 NHVALACMIIFLTIAYLKLLFLKKYS 147

DB 629 NHVALACMIVIFLTIAYLKLLFLKKYS 655

RESULT 10

ABP52127

ID ABP52127 standard; Protein; 655 AA.

AC ABP52127;

DT 10-OCT-2002 (first entry)

XX Homo sapiens ABC transporter ABCG2 protein SEQ ID NO:79.

XX ATP-binding cassette transporter; ABC transporter; modulation; D loop;
 KW cancer; bacterial infection; fungal infection; protozoal infection;
 KW antibacterial; fungicide; protozoicide.

OS Homo sapiens.

PN EPI217066-A1.

XX 26-JUN-2002.

XX 21-DEC-2000; 2000EP-0870316.

XX 21-DEC-2000; 2000EP-0870316.

XX (UYGE-) UNIV GENT.

XX WPI; 2002-550404/59.

XX Modulating activity of ATP-binding cassette (ABC) transporters by
 PT influencing dimerization of nucleotide binding domains through use of D
 PT loop sequence of an ABC transporter, or its antisense peptide or
 PT peptide mimetic -

XX Disclosure; Fig 3; 290pp; English.

XX The present invention describes a method (M1) for modulating the activity
 CC of ATP-binding cassette (ABC) transporters by influencing the
 CC dimerization of the nucleotide binding domains comprising: (a) a
 CC polypeptide (polyP) consisting of 5-50 amino acids comprising the D loop
 CC sequence of an ABC transporter (ABP52049 to ABP52091); (b) a polyP
 CC consisting of the D loop sequence of an ABC transporter; (c) a peptide
 CC mimetic or antisense peptide of (a) or (b). ABC transporters have

CC antibacterial, fungicide and protozoacide activities. (M1) is useful for
CC selectively modulating the activity of ABC transporters belonging to the
CC group of multidrug transporter/P-glycoproteins. Bacterial, fungal or
CC protozoal ABC transporters are involved in the infection of a mammal or
CC in the induction of resistance to antibiotics or drugs in a mammal. (M1)
CC is useful for preventing, treating or alleviating diseases associated
CC with functionality of an ABC transporter. ABP52092 to ABP52140 represent
CC ABC transporter proteins given in the exemplification of the present
CC invention.

XX
SQ Sequence 655 AA;
Query Match 81.0%; Score 623; DB 23; Length 655;
Best Local Similarity 81.6%; Pred. No. 1.5e-61;
Matches 120; Conservative 8; Mismatches 17; Indels 2; Gaps 1;
QY 1 FGLGAEAYTASSMALAIATGQSVSVATLLMTIAFVFNMLPSGLLVNLTIGPWLWLOQ 60
DB 511 FTLMMVAYSASSMALAIAGQSVSVATLLMTICFVFMIFSGLVNLTITIASWLSWLOQ 570
QY 61 FSPRYGFTALQYNEFLGQFCPGFNVDNSTCVNSYAICTGNEYLIHQIGIELSPWGLWK 120
DB 571 FSPRYGFTALQYNEFLGQFCPGFNVDNSTCVNSYAICTGNEYLIHQIGIELSPWGLWK 120
QY 121 NHVALACMIIFLTIAYLKLLFLKKYS 147
DB 629 NHVALACMIIVIFLTIAYLKLLFLKKYS 655

RESULT 11
AAU80028
ID AAU80028 standard; Protein; 655 AA.

XX AC AAU80028;
XX DT 15-JUL-2002 (first entry)
XX DE Human ABCG2.
XX KW Human; ABCG2; transporter protein; anticancer drug tolerance;
XX KW indocarbazole.
XX OS Homo sapiens.
XX PN WO200228894-A1.
XX PD 11-APR-2002.
XX PF 18-SEP-2001; 2001WO-JP08112.
XX PR 03-OCT-2000; 2000JP-0303441.
XX PA (BANY) BANYU PHARM CO LTD.
XX PI Komatani H, Hara Y, Kotani H, Nakagawa R;
XX DR WPI; 2002-352228/38.
XX DR N-PSDB; ABK49901.
XX PT ABCG2 gene encoding transporter protein capable of selectively
PT transporting indocarbazole compounds, useful in screening inhibitors
PT and anticancer agents for administration in chemotherapy -
XX
XX Claim 1; Page 71-76; 98pp; Japanese.

XX The invention relates to an ABCG2 gene encoding a transporter protein
CC capable of imparting tolerance to an anticancer agent in mammals
CC comprising a fully defined sequence as given in the specification or an
CC amino acid sequence based on the sequence but with some amino acids
CC substituted, deleted or added. The gene and encoded protein are useful
CC in screening inhibitors and anticancer agents for administration in
CC chemotherapy with enhancement in sensitivity of cancer cell tolerance.
CC The gene relating to drug tolerance can be modified e.g. with the

CC transporter inhibitors, screened compounds, antibodies and antisense
CC nucleotides. The transporter is capable of selectively transporting
CC indocarbazole compounds extracellularly. The present sequence
CC represents the amino acid sequence of human ABCG2 protein.

XX
SQ Sequence 655 AA;
Query Match 81.0%; Score 623; DB 23; Length 655;
Best Local Similarity 81.6%; Pred. No. 1.5e-61;
Matches 120; Conservative 8; Mismatches 17; Indels 2; Gaps 1;
QY 1 FGLGAEAYTASSMALAIATGQSVSVATLLMTIAFVFNMLPSGLLVNLTIGPWLWLOQ 60
DB 511 FTLMMVAYSASSMALAIAGQSVSVATLLMTICFVFMIFSGLVNLTITIASWLSWLOQ 570
QY 61 FSPRYGFTALQYNEFLGQFCPGFNVDNSTCVNSYAICTGNEYLIHQIGIELSPWGLWK 120
DB 571 FSPRYGFTALQYNEFLGQFCPGFNVDNSTCVNSYAICTGNEYLIHQIGIELSPWGLWK 120
QY 121 NHVALACMIIFLTIAYLKLLFLKKYS 147
DB 629 NHVALACMIIVIFLTIAYLKLLFLKKYS 655

RESULT 12
AAU80029
ID AAU80029 standard; Protein; 655 AA.

XX AC AAU80029;
XX DT 15-JUL-2002 (first entry)
XX DE Human ABCG2 mutant 482T.
XX KW Human; ABCG2; transporter protein; anticancer drug tolerance;
XX KW indocarbazole; mutant; mutein.
XX OS Homo sapiens.
XX OS Synthetic.
XX FH Key Location/Qualifiers
XX FT Misc-difference 482 /note= "Wild type Arg substituted by Thr"
XX FT
XX PN WO200228894-A1.
XX PD 11-APR-2002.
XX PF 18-SEP-2001; 2001WO-JP08112.
XX PR 03-OCT-2000; 2000JP-0303441.
XX PA (BANY) BANYU PHARM CO LTD.
XX PI Komatani H, Hara Y, Kotani H, Nakagawa R;
XX DR WPI; 2002-352228/38.
XX DR N-PSDB; ABK49911.
XX PT ABCG2 gene encoding transporter protein capable of selectively
PT transporting indocarbazole compounds, useful in screening inhibitors
PT and anticancer agents for administration in chemotherapy -
XX
XX Disclosure; Page 87-90; 98pp; Japanese.

XX The invention relates to an ABCG2 gene encoding a transporter protein
CC capable of imparting tolerance to an anticancer agent in mammals
CC comprising a fully defined sequence as given in the specification or an
CC amino acid sequence based on the sequence but with some amino acids
CC substituted, deleted or added. The gene and encoded protein are useful
CC in screening inhibitors and anticancer agents for administration in
CC chemotherapy with enhancement in sensitivity of cancer cell tolerance.
CC The gene relating to drug tolerance can be modified e.g. with the

CC transporter inhibitors, screened compounds, antibodies and antisense
CC nucleotides. The transporter is capable of selectively transporting
CC indocarbazole compounds extracellularly. The present sequence
CC represents the amino acid sequence of human ABCG2 mutant 482T.

XX SQ Sequence 655 AA;

Query Match 81.0%; Score 623; DB 23; Length 655;
Best Local Similarity 81.6%; Pred. No. 1.5e-61;
Matches 120; Conservative 8; Mismatches 17; Indels 2; Gaps 1;

QY 1 FGLGAEAYTASSMALAIATGQSVSVATLLMTIAFVFMFLFSGLLVNLRTIGPWLWLQY 60
DB 511 FTLMMVAYSASSMALAIAGQSVSVATLLMTICFVFMFLFSGLLVNLRTIASWLSWLO 570
QY 61 FSPRYGFTALQYNEFLGQFCFNVNTDSTCVNSYAICTGNEYVLNQIELSPWGLWK 120
DB 571 FSPRYGFTALQYNEFLGQFCFGLNATGNPC--NYATCTGEEYLVKQIDLSWGLWK 628
QY 121 NHVALACMIIFLTIAVLKLLFLKKYS 147
DB 629 NHVALACMIVIFLTIAVLKLLFLKKYS 655

RESULT 13

AAO14781
ID AAO14781 standard; Protein; 655 AA.

XX AC AAO14781;

XX DT 28-JUN-2002 (first entry)

XX DE Human BCRP protein.

XX KW Human; BCRP protein; membrane penetrating region; cancer.

XX OS Homo sapiens.

XX PN JP2002065277-A.

XX PD 05-MAR-2002.

XX PF 31-AUG-2000; 2000JP-0263742.

XX PR 31-AUG-2000; 2000JP-0263742.

XX PA (GANK-) ZH GAN KENYUKAI.

XX DR WPI; 2002-324198/36.

XX DR N-PSDB; AAL42412.

XX PT Mutant BCRP protein useful for treatment of cancer -

XX PS Claim 13; Page 7-8; 15pp; Japanese.

XX CC The invention comprises a mutant human BCRP protein, having a deletion,
CC replacement or addition of at least one amino acid in the fifth membrane
CC penetrating region of the wild-type BCRP protein. The mutant BCRP protein
CC can be used for the treatment of cancer. The present amino acid sequence
CC represents a human BCRP protein.

XX SQ Sequence 655 AA;

Query Match 81.0%; Score 623; DB 23; Length 655;
Best Local Similarity 81.6%; Pred. No. 1.5e-61;
Matches 120; Conservative 8; Mismatches 17; Indels 2; Gaps 1;

QY 1 FGLGAEAYTASSMALAIATGQSVSVATLLMTIAFVFMFLFSGLLVNLRTIGPWLWLQY 60
DB 511 FTLMMVAYSASSMALAIAGQSVSVATLLMTICFVFMFLFSGLLVNLRTIASWLSWLO 570
QY 61 FSPRYGFTALQYNEFLGQFCFNVNTDSTCVNSYAICTGNEYVLNQIELSPWGLWK 120
DB 571 FSPRYGFTALQYNEFLGQFCFGLNATGNPC--NYATCTGEEYLVKQIDLSWGLWK 628

DB 571 FSPRYGFTALQYNEFLGQFCFGLNATGNPC--NYATCTGEEYLVKQIDLSWGLWK 628
QY 121 NHVALACMIIFLTIAVLKLLFLKKYS 147
DB 629 NHVALACMIVIFLTIAVLKLLFLKKYS 655

RESULT 14

ABB07270
ID ABB07270 standard; Protein; 655 AA.

XX AC ABB07270;

XX DT 26-MAR-2002 (first entry)

XX DE Human BCRP (huBCRP) amino acid sequence.

XX KW Stem cell; ATP transport protein; ATP-binding cassette; antiparkinsonian;
KW hepatotropic; neurodegenerative; cytostatic; antianemic; muscular; BCRP;
KW cardiant; gene therapy.

XX OS Homo sapiens.

XX PN WO200192877-A2.

XX PD 06-DEC-2001.

XX PF 30-MAY-2001; 2001WO-US17459.

XX PR 31-MAY-2000; 2000US-0584586.

XX PR 29-MAY-2001; 2001US-0866866.

XX PA (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.

XX PI Sorrentino B, Schuetz J;

XX DR WPI; 2002-114368/15.

XX DR N-PSDB; ABA94369.

XX PT Identifying a stem cell, for treating e.g., muscular dystrophy,
PT myocardial infarction, Parkinson's disease, or neurodegenerative
PT disorders, comprises detecting the expression of an ATP transport
PT protein (BCRP) by a cell -

XX PS Examples; Page 78-80; 87pp; English.

XX CC The invention provides a method of identifying and/or isolating a stem
CC cell that involves detecting the expression of an ATP transport protein
CC containing a conserved ATP-binding cassette (BCRP) by a cell in a sample
CC comprising stem cells. The isolated stem cells may be used in the
CC treatment of diseases such as muscular dystrophy, degenerative liver
CC disorder, myocardial infarction, Parkinson's disease, degenerative
CC disorders of the brain, and for tissue regeneration or replacement.
CC Haematopoietic cells can be used in bone marrow transplants (e.g., for
CC treatment of leukemia) and for ex vivo gene therapy for treating blood
CC diseases such as sickle cell anemia and thalassemia. The stem cells can
CC also be used as cell targets in gene therapy protocols. The present
CC sequence represents the amino acid sequence of human BCRP.

XX SQ Sequence 655 AA;

Query Match 81.0%; Score 623; DB 23; Length 655;
Best Local Similarity 81.6%; Pred. No. 1.5e-61;
Matches 120; Conservative 8; Mismatches 17; Indels 2; Gaps 1;

QY 1 FGLGAEAYTASSMALAIATGQSVSVATLLMTIAFVFMFLFSGLLVNLRTIGPWLWLQY 60
DB 511 FTLMMVAYSASSMALAIAGQSVSVATLLMTICFVFMFLFSGLLVNLRTIASWLSWLO 570
QY 61 FSPRYGFTALQYNEFLGQFCFNVNTDSTCVNSYAICTGNEYVLNQIELSPWGLWK 120
DB 571 FSPRYGFTALQYNEFLGQFCFGLNATGNPC--NYATCTGEEYLVKQIDLSWGLWK 628

QY 121 NHVALACMIIIFLTAYLKLLFLKKYS 147
DB 629 NHVALACMIVIFLTAYLKLLFLKKYS 655

Db 629 NHVALACMIVIFLTAYLKLLFLKKYS 655

Search completed: August 1, 2003, 18:47:59
Job time : 22.3616 secs

RESULT 15

ABB07273
ID ABB07273 standard; Protein; 655 AA.

XX AC ABB07273;

XX DT 26-MAR-2002 (first entry)

XX DE Human BCRP (huBCRP) sequence.

XX KW Stem cell; ATP transport protein; ATP-binding cassette; antiparkinsonian;
KW hepatotropic; neurodegenerative; cytostatic; antianemic; muscular; BCRP;
KW cardiant; gene therapy.

XX OS Homo sapiens.

XX PN WO200192877-A2.

XX PD 06-DEC-2001.

XX PF 30-MAY-2001; 2001WO-US17459.

XX PR 31-MAY-2000; 2000US-0584586.

XX PR 29-MAY-2001; 2001US-0866866.

XX PA (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.

XX PI Sorrentino B, Schuetz J;

XX DR MPI; 2002-114368/15.

XX DR N-PSDB; ABA94383.

PT Identifying a stem cell, for treating e.g., muscular dystrophy,
PT myocardial infarction, Parkinson's disease, or neurodegenerative
PT disorders, comprises detecting the expression of an ATP transport
PT protein (BCRP) by a cell

XX Example 1; Page 86-87; 87pp; English.

PS The invention provides a method of identifying and/or isolating a stem
CC cell that involves detecting the expression of an ATP transport protein
CC containing a conserved ATP-binding cassette (BCRP) by a cell in a sample
CC comprising stem cells. The isolated stem cells may be used in the
CC treatment of diseases such as muscular dystrophy, degenerative liver
CC disorder, myocardial infarction, Parkinson's disease, degenerative liver
CC disorders of the brain, and for tissue regeneration or replacement.
CC Haematopoietic cells can be used in bone marrow transplants (e.g., for
CC treatment of leukemia) and for ex vivo gene therapy for treating blood
CC diseases such as sickle cell anemia and thalassemia. The stem cells can
CC also be used as cell targets in gene therapy protocols. The present
CC sequence represents the human BCRP.

XX SQ Sequence 655 AA;

Query Match 81.0%; Score 623; DB 23; Length 655;

Best Local Similarity 81.6%; Pred. No. 1.5e-61;

Matches 120; Conservative 8; Mismatches 17; Indels 2; Gaps 1;

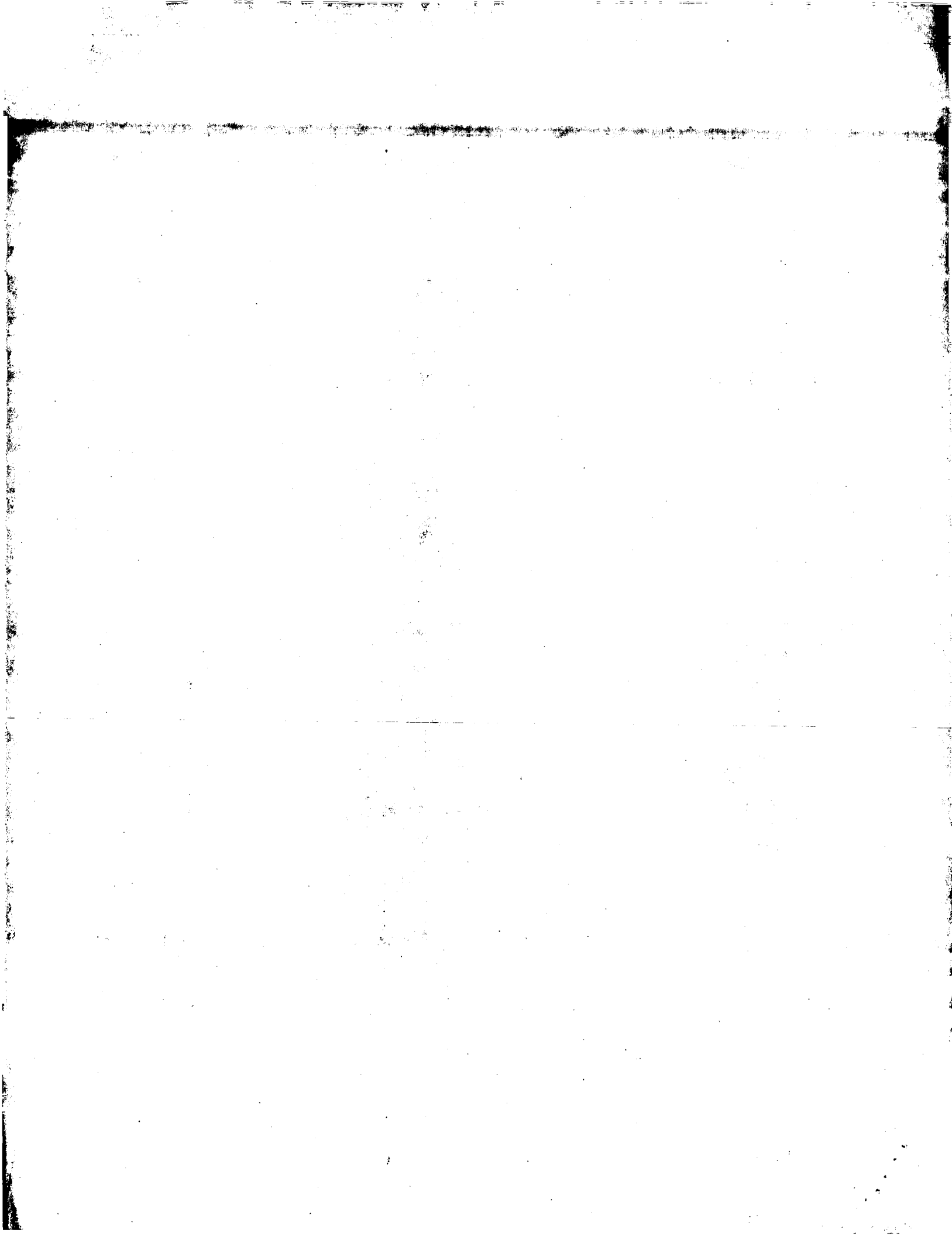
QY 1 FGLGAEAYTASSMALAATGQSVSVATLLMTIAFVEMWLFSGLLVNLRTIGPWLSQLQY 60

DB 511 FTLMMVAISASSMALAIAAGQSVSVATLLMTICFVFMWIFSGLLVNLRTIASWLSWQY 570

QY 61 FSPRYGFTALQYNBFLQGFQCFGNVTDNSTCVNSYAICTGNEYLINQIELSPWGLWK 120

DB 571 FSPRYGFTALQHNFLQGFQCFGLNATGNFNC--NYATCTGEEVLVKQIDLSWGLWK 628

QY 121 NHVALACMIIIFLTAYLKLLFLKKYS 147



GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model
Run on: August 1, 2003, 18:45:06 ; Search time 22.9115 Seconds
(without alignments)
1655.666 Million cell updates/sec

Title: US-09-856-927-4
Perfect score: 769
Sequence: 1 FGLGAEYTAGSMALAIATG.....MTIIFLTATYKLLFLKKYS 147

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_23.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	769	100.0	147	11 Q921T0	Q921T0 mus musculus
2	739	96.1	657	11 Q9R004	Q9R004 mus musculus
3	623	81.0	655	4 Q96TA8	Q96TA8 homo sapien
4	623	81.0	655	4 Q96LD6	Q96LD6 homo sapien
5	623	81.0	655	4 Q81X16	Q81X16 homo sapien
6	621	80.8	656	6 Q8MIB3	Q8MIB3 sus scrofa
7	415.5	54.0	650	11 Q8BK15	Q8BK15 mus musculus
8	146	19.0	1468	10 Q8GU85	Q8GU85 oryza sativ
9	143	18.6	801	5 Q8T691	Q8T691 dictyosteli
10	134	17.4	1315	10 Q8GU86	Q8GU86 oryza sativ
11	133	17.3	547	5 Q9U2D0	Q9U2D0 caenorhabdi
12	133	17.3	1435	10 Q9FZ13	Q9FZ13 arabidopsis
13	132.5	17.2	646	11 Q8K4E1	Q8K4E1 mus musculus
14	132.5	17.2	1619	3 O13407	O13407 magnaporthe
15	132	17.2	1501	3 Q96VK5	Q96VK5 emericeilla
16	131.5	17.1	646	11 Q8BXW2	Q8BXW2 mus musculus

17	128.5	16.7	423	11 Q8RIU5	Q8RIU5 mus musculu
18	128.5	16.7	627	11 Q91WA9	Q91WA9 mus musculu
19	128.5	16.7	646	11 Q8VBS9	Q8VBS9 mus musculu
20	128.5	16.7	1434	10 Q9C623	Q9C623 arabidopsis
21	128	16.6	687	5 Q94G60	Q94G60 drosophila
22	127.5	16.6	1501	3 Q96VL9	Q96VL9 botrytis ci
23	127	16.5	1426	3 Q96W59	Q96W59 mycosphaera
24	126	16.4	1439	5 Q8T679	Q8T679 dictyosteli
25	126	16.4	1450	10 Q9SJR6	Q9SJR6 arabidopsis
26	125.5	16.3	687	5 Q9NH94	Q9NH94 bombyx mori
27	123.5	16.1	416	4 Q9NT30	Q9NT30 homo sapien
28	123	16.0	1475	3 Q9P4V5	Q9P4V5 botrytis ci
29	122.5	15.9	708	10 Q9M2V5	Q9M2V5 arabidopsis
30	121.5	15.8	740	10 Q8O946	Q8O946 arabidopsis
31	121.5	15.8	1459	3 Q8TFM7	Q8TFM7 monilinia f
32	121	15.7	684	5 Q9N4Z3	Q9N4Z3 caenorhabdi
33	121	15.7	1466	3 P78576	P78576 emericeilla
34	118.5	15.4	1439	3 Q9UW03	Q9UW03 botrytis ci
35	118	15.3	1413	10 Q9ZUT8	Q9ZUT8 arabidopsis
36	117	15.2	785	4 Q96L76	Q96L76 homo sapien
37	116.5	15.1	1500	3 Q8NJ54	Q8NJ54 candida dub
38	116	15.1	669	5 Q8WR52	Q8WR52 tribolium c
39	116	15.1	669	5 Q8WR51	Q8WR51 tribolium c
40	116	15.1	1396	3 Q9C0Y5	Q9C0Y5 schizosacch
41	116	15.1	1441	5 Q8T681	Q8T681 dictyosteli
42	115.5	15.0	1491	3 Q96W99	Q96W99 gibberella
43	115	15.0	1499	3 Q9Y839	Q9Y839 mycosphaera
44	114.5	14.9	453	3 Q02922	Q02922 saccharomyc
45	114.5	14.9	662	10 Q9M2V6	Q9M2V6 arabidopsis

ALIGNMENTS

RESULT 1

ID	Q921T0	PRELIMINARY;	PRT;	147 AA.
AC	Q921T0;			
DC	01-MAY-1999 (Tremblrel. 10, Created)			
DT	01-MAY-1999 (Tremblrel. 10, Last sequence update)			
DT	01-DEC-2001 (Tremblrel. 19, Last annotation update)			
DE	Placenta-specific ATP-binding cassette transporter (Fragment).			
GN	ABCG2.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=Placenta;			
RX	MEDLINE=99065313; PubMed=9850061;			
RA	Allikmets R., Schriml L.M., Hutchinson A., Romano-Spica V., Dean M.;			
RT	"A human placenta-specific ATP-binding cassette gene (ABCP) on			
RL	chromosome 4q22 that is involved in multidrug resistance.";			
DR	EMBL; AF103875; AAD09189.1; -.			
DR	MGI; MGI:1347061; Abcg2.			
KW	ATP-binding.			
FT	NON_TER 1 1			
SQ	SEQUENCE 147 AA; 16355 MW; 93CD633CFF94889F CRC64;			

Query Match 100.0%; Score 769; DB 11; Length 147;
Best Local Similarity 100.0%; Pred. No. 4.5e-68;
Matches 147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	FGLGAEYTAGSMALAIATGQSVSVATLLMTIAFVPMFLSGLLVNRTTGPWLSWLOQ	60
Db	1	FGLGAEYTAGSMALAIATGQSVSVATLLMTIAFVPMFLSGLLVNRTTGPWLSWLOQ	60
QY	61	FSIPRYGPTALQYNEFLGQEFPCPGFNVTDNSTCVNSYAICTGNEYLINQIELSPWGLWK	120
Db	61	FSIPRYGPTALQYNEFLGQEFPCPGFNVTDNSTCVNSYAICTGNEYLINQIELSPWGLWK	120

```

QY 121 NHVALACMIIFLTITAYLKLLFLKKYS 147
DB 121 NHVALACMIIFLTITAYLKLLFLKKYS 147

RESULT 2
Q9R004
ID Q9R004 PRELIMINARY; PRT; 657 AA.
AC Q9R004;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-WAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Breast cancer resistance protein 1.
GN ABCG2 OR BCRP1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB; TISSUE=Liver;
RX MEDLINE=99413474; PubMed=10485464;
RA Allen J.D., Brinkhuis R.F., Wijnholds J., Schinkel A.H.;
RT "The mouse Bcrp1/Mxr/Abcg gene: amplification and overexpression in
RT cell lines selected for resistance to topotecan, mitoxantrone, or
RT doxorubicin.";
RL Cancer Res. 59:4237-4241(1999).
DR EMBL; AF140218; AAD54216.1; -.
DR MGD; MGI:1347061; Abcg2.
DR InterPro; IPR003593; AAA ATPase.
DR InterPro; IPR003439; ABC transporter.
DR InterPro; IPR006162; Ppantne_attach.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; 1.
KW ATP-binding.
SQ SEQUENCE 657 AA; 73021 MW; 207B70BC272CC0D5 CRC64;

Query Match 96.1%; Score 739; DB 11; Length 657;
Best Local Similarity 97.3%; Pred. No. 1.9e-64;
Matches 143; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 FGLGAEAYTASSMALAIATGQSVSVATLLMTIAFVFMFLFSGLLVNLRTIGPWLSQLQY 60
DB 511 FTLIMVAYSSASSMALAIATGQSVSVATLLMTIAFVFMFLFSGLLVNLRTIGPWLSQLQY 570

QY 61 FSPRYGFTALQYNEFLGQEFPCGFNVTDNSTCVNSYAICTGNEYLINQGIELSPWLK 120
DB 571 FSPRYGFTALQYNEFLGQEFPCGFNVTDNSTCVNSYAICTGNEYLINQGIELSPWLK 630

QY 121 NHVALACMIIFLTITAYLKLLFLKKYS 147
DB 631 NHVALACMIIFLTITAYLKLLFLKKYS 657

RESULT 3
Q96TA8
ID Q96TA8 PRELIMINARY; PRT; 655 AA.
AC Q96TA8;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-WAR-2003 (TrEMBLrel. 23, Last annotation update)
DE ATP-binding cassette superfamily G (White) member 2.
GN ABCG2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21201983; PubMed=11306452;
RA Komatani H., Kotani H., Hara Y., Nakagawa R., Matsumoto M.,

RA Arakawa H., Nishimura S.;
RT "Identification of breast cancer resistant protein/mitoxantrone
RT resistance/placenta-specific, ATP-binding cassette transporter as a
RT transporter of NB-506 and J-107089, topoisomerase I inhibitors with an
RT indolocarbazole structure.";
RL Cancer Res. 61:2827-2832(2001).
DR EMBL; AB051855; BAB46933.1; -.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR006162; Ppantne_attach.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transporter; 1.
DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; 1.
KW ATP-binding.
SQ SEQUENCE 655 AA; 72314 MW; A9AF66B96034C5A8 CRC64;

Query Match 81.0%; Score 623; DB 4; Length 655;
Best Local Similarity 81.6%; Pred. No. 4.9e-53;
Matches 120; Conservative 8; Mismatches 17; Indels 2; Gaps 1;

QY 1 FGLGAEAYTASSMALAIATGQSVSVATLLMTIAFVFMFLFSGLLVNLRTIGPWLSQLQY 60
DB 511 FTLIMVAYSSASSMALAIATGQSVSVATLLMTIAFVFMFLFSGLLVNLRTIGPWLSQLQY 570

QY 61 FSPRYGFTALQYNEFLGQEFPCGFNVTDNSTCVNSYAICTGNEYLINQGIELSPWLK 120
DB 571 FSPRYGFTALQYNEFLGQEFPCGFNVTDNSTCVNSYAICTGNEYLINQGIELSPWLK 628

RESULT 4
Q96LD6
ID Q96LD6 PRELIMINARY; PRT; 655 AA.
AC Q96LD6;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-WAR-2003 (TrEMBLrel. 23, Last annotation update)
DE ABC transporter ABCG2.
GN ABCG2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Schuetz J.D., Wall A.M., Sampath J., Sorrentino B., Du G.;
RT "The Human ABC Transporter, ABCG2, Transports Hoechst 33342 and
RT Requires an Intact Walker A Motif.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY017168; AAG52982.1; -.
DR InterPro; IPR003593; AAA ATPase.
DR InterPro; IPR003439; ABC transporter.
DR InterPro; IPR006162; Ppantne_attach.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; 1.
KW ATP-binding.
SQ SEQUENCE 655 AA; 72288 MW; B3B5DC02C095CA8 CRC64;

Query Match 81.0%; Score 623; DB 4; Length 655;
Best Local Similarity 81.6%; Pred. No. 4.9e-53;
Matches 120; Conservative 8; Mismatches 17; Indels 2; Gaps 1;

QY 1 FGLGAEAYTASSMALAIATGQSVSVATLLMTIAFVFMFLFSGLLVNLRTIGPWLSQLQY 60
DB 511 FTLIMVAYSSASSMALAIATGQSVSVATLLMTIAFVFMFLFSGLLVNLRTIGPWLSQLQY 570

QY 61 FSPRYGFTALQYNEFLGQEFPCGFNVTDNSTCVNSYAICTGNEYLINQGIELSPWLK 120
DB 571 FSPRYGFTALQYNEFLGQEFPCGFNVTDNSTCVNSYAICTGNEYLINQGIELSPWLK 628

```

```
QY 121 NHVALACMIIIFLTAYLKLFLKYS 147
DB 629 NHVALACMIVIFLTAYLKLFLKYS 655

RESULT 5
Q8IX16
ID Q8IX16 PRELIMINARY; PRT; 655 AA.
AC Q8IX16;
DT 01-MAR-2003 (TremBLrel. 23, Created)
DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)
DE ATP-binding cassette protein ABCG2.
GN ABCG2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Yoshikawa M., Yabuuchi H., Ikegami Y., Ishikawa T.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF463519; AAO14617.1; -.
KW ATP-binding.
SQ SEQUENCE 655 AA; 72314 MW; ABAF60B591D4C5A8 CRC64;

Query Match 81.0%; Score 623; DB 4; Length 655;
Best Local Similarity 81.6%; Pred. No. 4.9e-53;
Matches 120; Conservative 8; Mismatches 17; Indels 2; Gaps 1;

QY 1 FGLGAEYATASSMALAIATGQSVSVATLLMTIAFVFMMLFSGLLVNLRTIGPWLWSLQY 60
DB 511 FTLMMVAYSASSMALAIAGQSVSVATLLMTICFVFMIFSGLLVNLRTIASWLSLQY 570

QY 61 FSPRYGFTALQYNEFLGQFCPGFNVTDNSTCVNSYAICTGNEYLINQIGIELSPWGLWK 120
DB 571 FSPRYGFTALQYNEFLGQFCPGFNVTDNSTCVNSYAICTGNEYLINQIGIELSPWGLWK 628

QY 121 NHVALACMIIIFLTAYLKLFLKYS 147
DB 629 NHVALACMIVIFLTAYLKLFLKYS 655

RESULT 6
Q8MIB3
ID Q8MIB3 PRELIMINARY; PRT; 656 AA.
AC Q8MIB3;
DT 01-OCT-2002 (TremBLrel. 22, Created)
DT 01-OCT-2002 (TremBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)
DE Brain multidrug resistance protein.
GN BMDP.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RA Eisenlaetter T., Galla H.J.;
RL "A new multidrug resistance protein at the blood-brain barrier.";
RL Biochem. Biophys. Res. Commun. 293:1273-1278 (2002).
DR EMBL: AJ420927; CAD12785.1; -.
DR InterPro: IPR003593; AAA_ATPase.
DR InterPro: IPR003439; ABC_transporter.
DR InterPro: IPR006162; Ppantne_attach.
DR Pfam: PF00005; ABC_tran; 1.
DR ProDom: PD000006; ABC_transporter; 1.
DR SMART: SM00382; AAA; 1.
DR PROSITE: PS00012; PHOSPHOTANTHETINE; 1.
KW ATP-binding.
SQ SEQUENCE 656 AA; 72392 MW; 118ADD5B53D9D67F CRC64;

Query Match 80.8%; Score 621; DB 6; Length 656;
Best Local Similarity 81.6%; Pred. No. 7.8e-53;
Matches 120; Conservative 11; Mismatches 14; Indels 2; Gaps 1;

QY 1 FGLGAEYATASSMALAIATGQSVSVATLLMTIAFVFMMLFSGLLVNLRTIGPWLWSLQY 60
DB 512 FTLMMVAYSASSMALAIAGQSVSVATLLMTISFVFMIFSGLLVNLRTIVVPLWSLQY 571

QY 61 FSPRYGFTALQYNEFLGQFCPGFNVTDNSTCVNSYAICTGNEYLINQIGIELSPWGLWK 120
DB 572 FSPRYGFTALQYNEFLGQFCPGFNVTDNSTCVNSYAICTGNEYLINQIGIELSPWGLWK 629

QY 121 NHVALACMIIIFLTAYLKLFLKYS 147
DB 630 NHVALACMIVIFLTAYLKLFLKYS 656

RESULT 7
Q8BKIS
ID Q8BKIS PRELIMINARY; PRT; 650 AA.
AC Q8BKIS;
DT 01-MAR-2003 (TremBLrel. 23, Created)
DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)
DE ATP-binding cassette.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=C57BL/6J; TISSUE=EYE;
RX MEDLINE=22354683; PubMed=12466851;
RA The PANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
DR EMBL: AK051880; BAC34799.1; -.
SQ SEQUENCE 650 AA; 73613 MW; B9995BDBDFD945E1 CRC64;

Query Match 54.0%; Score 415.5; DB 11; Length 650;
Best Local Similarity 58.0%; Pred. No. 1.3e-32;
Matches 80; Conservative 22; Mismatches 35; Indels 1; Gaps 1;

QY 7 AYATASSMALAIATGQSVSVATLLMTIAFVFMMLFSGLLVNLRTIGPWLWSLQY 66
DB 509 AYSASSLPLSIGAGENAVAVPTLLVTIYFVFMPLFSGLSLGSFPLKLSWIQYPSIPHY 568

QY 67 GFTALQYNEFLGQFCPGFNVTDNSTCVNSYAICTGNEYLINQIGIELSPWGLWKHVALA 126
DB 569 GFRALLHNEFLGQFCPEHNTTEVSRCN-YVICTGEEFLMIQIGIDLSSGFWENHLALV 627

QY 127 CMIIIFLTAYLKLFLK 144
DB 628 CTMIILLTITVYQLQVK 645

RESULT 8
Q8GU85
ID Q8GU85 PRELIMINARY; PRT; 1468 AA.
AC Q8GU85;
DT 01-MAR-2003 (TremBLrel. 23, Created)
DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)
DE PDR-like ABC transporter.
GN PDR4.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzae; Oryza.
OX NCBI_TaxID=39947;
```


RESULT 13

